

Means to public understanding

The British Association is hitching its wagon to the campaign for public understanding of science. It should be cautious in its manner but more daring in its advocacy of causes.

NEXT week's meeting of the British Association for the Advancement of Science, generally known in Britain as "the BA", promises well — as these occasions go. The location (Oxford) is attractive, there is a greater awareness among the British public that the future is linked with what goes on in laboratories, although just how is far from clear, while active researchers, conspicuous absentees at annual meetings of this kind, are now in Britain half-convinced that their present difficulties stem from past neglect of public representation. It is also likely that this year's president (Sir Walter Bodmer), from the generation of researchers still untarnished by the consequences of past mistakes, can be counted on for a rousing speech.

That is the good news. The other side of the coin is that it is still not clear how the BA sees its role in contemporary society. That is no great shame. Most similarly named organizations scattered about the English-speaking world are similarly at a loss. The Australia and New Zealand association (ANZAAS) seems to be an annual disappointment. The annual Indian Science Congress generates enthusiasm, but fitfully. Only the US association (AAAS), anchored firmly in a successful and admirable publishing business, has been able to perform a wider and continuing function, in matters as different as the training of scientists as journalists, the freedom of scientists in repressive regimes and the annual scrutiny of the federal budget. What has gone wrong? And why?

The history is familiar, but worth repeating. The BA may seem a nineteenth-century creation (midway between the reform bills of 1830 and 1832 that enfranchised Britain), but in reality its roots lie in the late eighteenth century and the then fashionable doctrine of Improvement by Taking Thought. The Josiah Wedgwoods and Joseph Priestleys of the time were a novel breed.

Their enterprise may have helped to make British agriculture efficient and to found the industrial revolution, but their intellectual curiosity helped to provoke subversive questions in many people's minds, making Britain a more interesting place. At the BA's foundation, Faraday had only just begun on his great life's work, Britain was rich in ingenious engineers but science was entirely unprofessionalized. For the remainder of the nineteenth century and a little beyond it, the annual meetings of the BA were occasions at which Britain's growing army of interested souls could satisfy their intellectual curiosity. Now they are occasions when some of those who know what happens in laboratories instruct a few of those who do not.

One of the BA's objectives now is to be a part of a national effort, in Britain, to give people in general more awareness of what science is and what it can accomplish. The issue has been in the air for the past two years, since the Royal Society published the report of a study (of which Bodmer was the chairman) arguing the case for a wider public understanding of science. The virtues of this cause are plain. The world does need to understand, and to participate in, the technical decisions made on its behalf. More public understanding would mean more rational decisions, perhaps even less irrationality, and so on. The case as put rests only marginally on the belief that public understanding would beget bigger budgets for research, which is only proper.

To say that there are snags is not to ask that the BA (or any

other association) should abandon these causes, but is merely to be cautious. The problem, in contemporary Britain, is not so much ignorance of science but disaffection from it. Gone is the excitement, a century ago, at the discovery that evolution could illuminate people's understanding of their place in the scheme of things, half a century ago that the Universe is what it is and, more recently, at the possibility that electricity generated by nuclear fusion would be so cheap that the costs of metering it would be an encumbrance. Developments such as these served not merely to fire the general imagination and to attract able young men and women to an honourable profession; they also augmented the general standing of intellectual life. The fact that the last of these three promises proved false is only part of the reason for present disaffections: it is more relevant that the once clear prospects offered by discovery have been muddled in the proper wish to make more deliberate use of it.

One pitfall for the campaign for the public understanding of science is that it is a means by which people in white coats risk seeming to patronize other grown-ups. So far, much attention has been paid to fuller and more expert explanation. But understanding as such is not the central issue. It matters more that the intellectual attention and imagination of adults, young and old, are not being sufficiently engaged by what at present passes for understanding. One symptom of this disengagement is the decline of the numbers of those making careers in science, concealing a no doubt disproportionate decline in the numbers of those of high quality who are thus inclined. That trends of the same kind are mirrored in comparable countries in Western Europe and North America but also in the Soviet Union does not rid the British campaign (now being echoed by ANZAAS and AAAS) of the obligation of trying to make science intellectually interesting, as in the eighteenth century. Making it clear will not suffice.

The other obvious pitfall is the assumption that the process of discovery is value-free. Nothing could be further from the truth. The process is sustained by intellectual curiosity, the belief that all discoveries are worthwhile and the conviction that their practical applications, if properly managed, are beneficial. The danger is that, after two decades of seemingly endless argument (which will continue) about the utility or otherwise of particular discoveries, the world is full of potential helpers in the campaign for understanding who behave as if discovery is principally a social nuisance.

That is why the BA should seek to awaken public interest as a means of engendering public understanding, which is most effectively done by taking public positions on matters important to research, the engine of discovery. While the BA has commendably, in recent years, offered its annual meeting as a forum for those wishing to discuss contentious (but not over-contentious) issues, it does not itself advocate much of importance.

British research has shamefully declined without protest from the BA, which has no opinion on the prospect of restriction on embryo research or even on the causes of the transition from the Cretaceous to the Tertiary. The robust figures populating the age of improvement behaved differently. Times have changed, but not that much. □



Still no government reply to Canada's AIDS report

- Epidemic only two or three years behind US
- Call for Cdn\$35 million per year on research

Washington

THIS week the number of people with AIDS in Canada is expected to pass 2,000, giving it a per capita incidence of AIDS amongst the highest in the developed world, less than the United States and on a par with France (see below). But despite the severity of the problem, the Canadian government has not yet responded to the recommendations of a report on AIDS issued by the Royal Society of Canada over three months ago.

The report, entitled *AIDS: A Perspective for Canadians*, is summarized in a 30-page document and supported by a 400-page volume of background papers. It was begun in May of 1987 at the behest of the Canadian Department of Health and Welfare, which paid for the bulk of its cost through Canada's Federal Center for AIDS. Most of the recommendations do not come with price tags attached, but one of the exceptions is the area of basic research into AIDS: the Royal Society calls for Cdn\$35 million to be set aside per year on AIDS research, even though only Cdn\$3.5 million is being spent now. The Royal Society further says that the research efforts should include social and behavioural studies as well as biological and medical research, and should be coordinated by a national committee.

The chairman of the committee that prepared the report, Michel Chrétien, says the AIDS epidemic in Canada lags behind that of the United States by about two or three years, and that roughly 70 per cent of the cases are located in the cities of Vancouver, Montreal and Toronto, where one-fourth of the population lives. The distribution of cases among the population is also markedly different from that in the United States and most of Europe: over 80 per cent of those affected are homosexual or bisexual men, and less than one per cent are heterosexual intravenous drug abusers.

The report takes a liberal stance, opposing mandatory testing for HIV infection, and supporting anti-discrimination legislation, education on condom use, and dispensing clean needles to drug addicts to prevent the spread of infection. Even so, many of the recommendations are vague, in recognition of the fact that the provincial governments have control over healthcare expenditures.

The report stresses the "economics of prevention", and gives equal attention to the special healthcare needs posed by the

AIDS epidemic, and to education and prevention activities.

Thirty-thousand Canadians are estimated to be infected with the human immunodeficiency virus (HIV), the virus causing AIDS, and eleven-thousand are projected to become ill by 1992, with their treatment taking up about 2 per cent of the nation's total health budget. The report

recommends investing \$80 million per year on AIDS education projects, and states that if the epidemic can be prevented from reaching US proportions, it will save Canada between Cdn\$2.5 and Cdn\$5 thousand million.

Alastair Clayton, the Director General of Canada's Federal Center for AIDS, says that his agency plans to spend \$52 million on AIDS education and prevention efforts out of its five-year budget of Cdn\$168 million. Clayton, who is drafting the government's official response to the report, identifies public education as a top priority; surveys reveal that 30–60 per cent of Canadians still believe that AIDS can be transmitted through environmental surfaces.

Carol Ezzell

■ AIDS testing in the newborn, page 7.

French government recognizes need for AIDS health campaign

Paris

THE new French Health Minister, Claude Evin, has returned to his desk after the summer recess determined to intensify the nation's efforts to combat the spread of AIDS. Evin is alarmed by recent analyses that predict, if current growth rates are maintained, as many as 21,000 AIDS sufferers in mainland France by the end of 1989. With 3,628 cases at the end of March, France has perhaps the highest per capita incidence of AIDS outside the United States.

A cornerstone of Evin's change of pace is the appointment of a well-known public health specialist, Claude Got, to provide an in-depth assessment of AIDS in France by the end of September. Got is professor of anatomy and pathology at the Ambroise-Paré hospital in the western suburbs of Paris. Over the past ten years he has played a leading role in various government public health enquiries, gaining a reputation for his hard-headed approach. Last year he resigned from a High Commission on alcoholism when the minister for communications authorized beer advertisements on television despite demands that they should be banned.

The prominent role of French scientists in identifying the human immunodeficiency virus (HIV) undoubtedly contributed to a false hope that, by pouring money into medical research, the disease would be conquered. Now, says Evin, "it is obvious that we can no longer be content to wait for a vaccine or effective treatment to be available in order to confront the disease". Although the previous health minister, Michèle Barzach, has been praised for her measures to combat AIDS, her information campaigns were understated and have had little apparent effect on behaviour. In his remit, Evin asks Got to

take a critical look at AIDS epidemiology, with the specific aim of guiding future public health campaigns.

Got's appointment coincides with the publication by the health department (Direction Générale de la Santé) of a demographic analysis of the known AIDS cases in France between 1978 and 1987. This study* shows that AIDS affects all socio-economic sectors of the French population, but that the geographical distribution of the disease, as well as its principal means of transmission, is changing.

While the Greater Paris area still has the highest proportion of AIDS sufferers (190 per million inhabitants), of which 73 per cent are homosexual and 7 per cent intravenous (i.v.) drug-users, this proportion of the national total has declined from 78 per cent in the late 1970s and early 1980s to 53 per cent in 1987. This decline is probably due, say the authors of the study, to the increasing importance of needle sharing by drug-users as a means of transmitting disease. In south-east France, where drug abuse is a growing problem, the number of AIDS sufferers (125 per million inhabitants) has increased from 6 per cent of the national total in 1978–84 to 17 per cent in 1987. In this region, 35 per cent of AIDS sufferers are drug-users and 41 per cent homosexual.

With i.v. drug-users clearly emerging as the population in which AIDS is spreading most rapidly, Claude Evin may, after all, have to take action which last month he considered too controversial for the new government. His minister delegate for health was asked to resign just 9 days after being appointed precisely because of his radical proposals to control drug abuse and the spread of AIDS (see *Nature* 334, 461; 1988).

Peter Coles

* *Bulletin épidémiologique hebdomadaire*, 8 August 1988.

Species jump may be responsible for seals' virus infection

London

THE virus that has killed several thousand common (or harbour) seals from the Baltic and North Seas during the past six months (see *Nature* 334, 553; 1988) is indistinguishable from canine distemper virus (CDV) the cause of the common infection of household pets and working dogs.

This astonishing finding is described in this week's *Nature* (Scientific Correspondence, page 20) by Dr Albert Osterhaus of the National Institute of Public Health and Environmental Protection in the Netherlands, and Dr Lies Vedder of the Pieterburen Seal Orphanage. The discovery raises the possibility that the virus has been transferred from dogs to seals in the recent past, although further investigations will be needed to prove that point.

Already there is evidence that the virus has spread from common seals to grey seals, and the Dutch finding is bound to raise concern for sea mammals other than seals, including dolphins and whales. At present there is no evidence that other sea mammals have been infected; but if the transfer of CDV to seals proves to have been recent, there may not yet have been time for that to have happened. Though the virus is not yet proven to be CDV, it is a very similar member of the same family of viruses, known as morbilliviruses, which includes those responsible for human measles and and rhinderpest in cattle, as well as a similar disease in goats.

An urgent programme to make a vaccine against the seal virus is now under way. The Dutch group has made about 100 doses of a possible vaccine that will be administered to seals in orphanages and sanctuaries in The Netherlands and Scandinavia this week, to tell whether the preparation offers protection to uninfected seals. But attempts, at the end of last week, to arrange for the production of materials for a vaccine in larger quantities through Coopers, the veterinary subsidiary of the Wellcome Foundation, have not yet materialized.

Osterhaus stresses that the usefulness of a vaccine may be limited. Even if the vaccine does prevent infection, there will be formidable problems in using it to protect wild animals. Why not use the vaccine against CDV which have been used to protect dogs? Osterhaus believes it would be wrong to use it in seals because of the risks of using a live-virus vaccine in a population of wild animals.

The Dutch claim that a CDV-like virus is responsible for the seal deaths of 1988 will surprise many virologists. Dr Cambell Cornwell at the Veterinary School at Glasgow, who is an expert on CDV, says that the reports, "if true, are sensational".

How a virus such as CDV can go "from dogs to pinnipeds makes the mind boggle". He says it will take some time to convince all researchers that CDV is the cause of the seal epidemic.

But the Dutch researchers believe there is evidence other than the properties of the virus to suggest that CDV has transferred from dogs to seals.

Dr Dernt Klingeborn, a virologist at the Uppsala centre, says that, this season, harp seals in Greenland waters migrated unusually far south, towards the coast of Denmark. He links this with an outbreak of CDV infection that wiped out 1,000 sleigh dogs in the very North of Greenland earlier this year, saying, "this is a very interesting coincidence".

Klingeborn believes the harp seals may have carried the virus to infect common seals in the North Sea and the Baltic, and now plans to check them for antibodies against the virus. He also says that the dogs in Greenland may have caught distemper by eating infected seal-meat. But Hans Jacob Helms of the Greenland administration believes the outbreak of distemper was imported from Canada either by fox or ravens, known carriers of CDV which also infects bears and wolves.

The effect of this year's epidemic on the future of the seal population is not yet clear. Osterhaus believes that this year's "wave of infection" will eventually "work itself out" as seals either become immune or succumb, but that "it'll probably come back", possibly doing less damage because some seals will then be immune.

Looking back on the false trails of the herpes and picornaviruses that misled him and his colleagues, Osterhaus says he now believes that these infections are consequences, not causes, of the seal disease.

There are no indications that people are susceptible to the seal version of the CDV virus. The researchers in close contact with infected seals say tests on themselves have all proved negative. Nor is there a danger that the seals have been infected by human measles virus, which has been eliminated as a cause in the course of the Dutch research.

Although the new research shows no direct link between seal disease and pollution, the possibility remains that pollution may play an important part in making seals more susceptible to infection and also by hampering the recovery of the seal population from this year's epidemic.

In collaboration with researchers in Britain, Osterhaus hopes to embark immediately on a £150,000 programme, with funds from the European Commission, to investigate links between pollution and the seal disease. **Christine McGourty**

Shuttle in September

Washington

THE launch date for the US space shuttle *Discovery* is now expected to be the last week of September, but National Aeronautics and Space Administration (NASA) officials are not prepared to set a precise date until a flight readiness review scheduled for 13 and 14 September, has been completed.

Meanwhile NASA and law enforcement officials have been investigating an apparent case of sabotage at the Hydra Pak plant in Utah where the shuttle's solid rocket booster O-rings are manufactured. Inspections revealed that several of the O-rings had been slashed with a razor or similar instrument, but NASA says that none of the damaged O-rings ever left the manufacturers plant. **J.P.**

Material objections

London

THE restructuring of the Science and Engineering Council (SERC) now under way resulted last month in the formation of a new Materials Commission which absorbs among other things a significant part of the responsibilities of the subcommittee for inorganic chemistry. Now researchers in chemistry are worried that inorganic chemistry might cease to exist as a separate and identifiable entity. In a letter last week to the head of SERC, Professor Bill Mitchell, 10 professors of chemistry, led by Professor Peter Maitlis of the University of Sheffield, criticize a rumoured move in this direction and request denial that such a move is being considered. Without issuing a straight denial, Mitchell says he will "try to ensure" that inorganic chemistry does not lose its identity. He says that absorption of the committee's responsibilities by the commission was a recognition of the subject's importance, and that funding for those areas taken over may increase. **C.McG.**

HERA here

Munich

THE 30 GeV electron ring of the HERA electron-proton collider at DESY (Deutsches Elektronen-Synchrotron), Hamburg, West Germany, was successfully run with its first electrons on 20 August. The test of the electron ring, at 7 GeV, showed that the pre-accelerator and central control facilities were working well.

Superconducting magnets for the PETRA ring where protons will be accelerated to 40 GeV before injection into the collider will be installed this autumn. Both rings will be ready for the first experiments in mid-1990.

HERA (Hadron Electron Ring Anlage) will be the world's first high-energy electron-proton collider. The ring, which is 6,336 metres long, will cost an estimated \$500 million. **S.D.**

Big payoff from high-temperature superconductor patents

Tokyo

DUPONT Company last week paid the University of Houston \$1.5 million dollars for exclusive rights to Professor Paul Chu's high-temperature superconductor patent applications. The company is betting that Chu's group will win the race for patent rights to the yttrium-based copper oxides discovered by Chu last year. If and when Chu receives the patent, DuPont will add another \$3 million to the payment.

A survey of patent applications for the new superconductors being constantly updated by Christopher Vear of the British Technology Group (see below) suggests that there will be some complex legal battles over patent priority and coverage.

In Europe and Japan, patent applications are published 18 months after filing (or after the earliest priority date claimed) and so can be monitored. US patents, on the other hand, remain secret until granted and are harder to track. But many of the published European and Japanese patent applications include priority claims for the key US patents and so give away US filing dates.

Vear's survey reveals that, among the patent applications now public, AT & T was first to file an application for oxide superconductors and devices made from them on January 9, 1987, followed three

yttrium-barium-copper oxide, he says. Chu will probably argue that the substitution of yttrium was "inventive" and led to a large increase in critical temperature.

IBM would appear to have missed the boat with its late patent application on January 23, 1987 for superconductors belonging to the lanthanum-barium-copper oxide group. That certainly seems the case in Japan and Europe where the "first to file" rule applies to patent priority. But IBM may still get the US patent.

The United States gives priority to the "first to invent" in a rule which usually only applies to research carried out within the country. But, according to Vear, if IBM Zurich transferred laboratory notes to any of IBM's US laboratories and the notes were understood and experiments repeated, the invention would be "reduced to practice" in the United States and IBM can rely on Bednorz and Müller's October 1986 paper on lanthanum-barium-copper oxide for priority there. But the extent to which such a patent could be extrapolated to cover other oxide superconductors is an open question.

The weakest patent of all appears to be that filed by Tokyo University on 17 January 1987. The patent only covers one form of lanthanum-barium-copper oxide. And, even if granted, would probably be

High-temperature superconductor patent applications

Organization	Priority claimed	Countries filed	Coverage
AT & T	9.1.87 3.3.87 10.3.87	Europe, USA	Two applications on a wide range of materials, methods and devices.
University of Houston	12.1.87 27.1.87 6.2.87 26.3.87	38 countries including USA, Europe, and Japan	Wide range of materials and production methods.
University of Tokyo	17.1.87	Japan	Relates to single-phase LaBCO only.
IBM	23.1.87	Europe, USA, Japan	Materials with LaBCO-type structure and stoichiometric composition and methods of fabrication.
MITI	27.1.87 20.2.87 5.3.87 11.6.87 30.8.87	12 countries including Japan, Europe and USA	Range of YBCO-type materials, including claimed room-temperature superconductor, and methods of preparation.

Table compiled by Dr C. Vear of British Technology Group.

days later by the University of Houston. These applications, however, relate to the lanthanum-barium-copper oxides discovered by Georg Bednorz and Alex Müller of IBM Zurich in 1986; yttrium-barium-copper oxide was not found by Chu until late January 1987 and he filed another patent application at that time.

The differences in stoichiometry, crystal structure and superconducting properties of these two types of oxide superconductor are likely to feature prominently in disputes over patent priority and coverage. Vear expects AT&T to argue that substitution of yttrium was "obvious". AT&T's first US application does include the possibility of a form of

of no commercial value. Japan's Ministry of International Trade and Industry on the other hand, has filed patents in 12 countries for yttrium-barium-copper oxide materials, including a claimed room-temperature superconductor, and the priority dates stretch back to 27 January 1987.

Despite the uncertainty, Chu's group is clearly confident of getting the patents for the yttrium-based superconductors. The group has filed patent applications in 38 countries, including some in Africa. And Julie Norris, a spokeswoman for the group, says they believe they are in a "strong position", otherwise "they would not be doing what they are doing".

David Swinbanks

Dear Bush and Dukakis...

Washington

LEADERS of US scientific societies representing more than three-quarters of a million scientists and engineers released a letter last week calling on the two presidential candidates to support "a strengthened and coherent science policy" in a new administration.

The letter stresses that problems "ranging from the destruction of the ozone layer ... to the spread of AIDS and restoration of the competitiveness of US industry" rely on scientific and technological resources. Leadership, the letter says must come from the White House.

The letter comes from 23 scientific societies, ranging from the Acoustical Society of America to the American Vacuum Society. The leaders of three of the biggest groups — Val Fitch, President of the American Physical Society (APS), Howard Schachman, president of the Federation of American Societies for Experimental Biology, and Russell Drew, president of the Institute of Electrical and Electronics Engineers and governor of the American Association of Engineering Societies — told a news conference that they were asking for a strong science advisor to be appointed early in the next administration.

The three society chiefs all spoke highly of the Eisenhower and Kennedy eras, when science had had a strong voice through the Presidential Science Advisory Committee, a body scrapped by President Nixon.

Fitch, a Nobel Prize winning physicist from Princeton, pointed out that the government "should not have to rely on an independent society" for advice on such matters as directed-energy weapons, where APS was the first to provide authoritative analysis.

Drew, president of Viking Instrument Corporation, argued that Congress had been forced to do the the administration's job in such legislation as the Omnibus Trade and Competitiveness Act. Among other measures, the act renames (from last week) the National Bureau of Standards as the National Institute of Technology and Standards and gives it new responsibilities for regional advanced technology outreach programmes. "A strong science and technology advisor will be needed to make them effective", said Drew.

The call for sounder advice seems unlikely to have much impact on the presidential race. A spokesman for the Republican party said that they were unlikely to send a formal reply to the letter as they receive "hundreds of letters like this".

Alun Anderson

Radioastronomers make plans to protect Antarctica and the Moon

Washington

RADIOASTRONOMERS, worried that many of their telescopes may be blinded by interference from other users of the electromagnetic spectrum, are planning early action to keep Antarctica, the far side of the Moon and deep space in their present pristine condition.

At last week's International Astronomical Union meeting in Washington, DC, resolutions were passed calling for international regulations to preserve certain parts of the radio spectrum for scientific use alone and to ensure that places so far unaffected by growing civil and military radio technologies remain unspoiled.

Discussing the triple threat posed to astronomy by urban lighting, radio interference and debris in low Earth orbit, astronomers noted some success on the first and last issues. In their efforts to keep the night sky dark and space clear of junk, astronomers find few opponents. But in the web of international rules and committees that must be negotiated to obtain allocations of radio frequency usage, they are fighting an increasing number of competitors.

Frequency bands are allocated to various use categories, of which radio-astronomy is one, by the International Telecommunications Union (ITU). ITU divides the world into three regions: the Americas; Europe, Africa and the Soviet Union; and Asia. But the biggest threat to radioastronomy is from global communication and navigation satellite networks, about which the old regional regulations say little, and from which Antarctica is not protected at all. So far, the South Pole is fairly free of radio interference, because few communication satellites operate in polar orbits, but astronomers want international regulations enacted to protect it in advance.

Around the rest of the world they are fighting a rearguard action now that all regions of the radio spectrum are under increasing pressure, and the frequency bands which used to be theirs by default have been encroached upon.

Although astronomers have few enemies, many of their competitors, which include international communications companies as well as various national governments, have economic and legislative power. The scientific community must therefore rely principally on persuasion. The Washington conference passed several resolutions asking for individual members and sympathetic governments to transmit astronomers' wishes to the International Radio Consultative Committee, which advises ITU on the competing needs of all radio users.

The conference also looked further into the future, at the feasibility of radio-astronomy being done from the far side of the Moon or from the Lagrangian point L-2, a position of orbital stability beyond the Moon, which Soviet scientists have mentioned as an ideal site for a space-based radio telescope. Until recently, anywhere beyond a line drawn through the centre of the Moon was considered deep space, and was heavily protected against radio use apart from frequency allocations for interplanetary probes. But recently the boundary of deep space was moved much further out, to two million km from Earth, bringing both the Moon and the L-2 point under the control of the same regulations that apply to terrestrial and satellite transmissions. Radioastronomers would like to shrink the size of near-Earth space back to its old boundaries before the Moon too is taken away from them.

David Lindley

*International Astronomical Union, Colloquium 112 on "Light pollution, radio interference and space debris", 13-16 August 1988.

Mismanagement charges at Stanford

Berkeley

AN internal investigation of health and safety issues at Stanford University found "many instances of mismanagement and waste", according to David Fetterman, who wrote the report of the investigation. A resignation letter written last December by a high-ranking employee to Stanford University's president Donald Kennedy, charging the university with numerous health and safety violations, gave rise to the investigation (see *Nature* 331, 382; 1988).

Fetterman's report also found that the critical employee, who requested anony-

mity, had been threatened with dismissal if he made his complaints known to the Stanford press office.

The report was particularly critical of a new incinerator, part of a new \$8-million Environmental Safety Facility. Fetterman says administrative tensions between the facilities management services and health and safety personnel have contributed to Stanford's safety problems.

Kennedy acknowledged that the report contained "harsh criticism", but added that the decision to make the report public would contribute to finding "constructive solutions".

Joseph Palca

Active optics comes of age

Munich

THE European Southern Observatory (ESO) has announced this month the successful testing of the 3.58-m primary mirror for its New Technology Telescope (NTT), which is to go into operation in Chile early next year. The mirror, the first of its kind to be used in a telescope, relies on so-called 'active optics' to monitor and optimize the images continuously. The image quality a threefold improvement upon that of previous telescopes.

Researchers at ESO spent ten years developing active optics, which use 75 mobile and 3 fixed supports beneath the mirror to control its shape. A micro-computer running a special algorithm is all that is required to control the feedback loop between the mirror and an image analyser. Only by slightly bending the



ESO's see-by-wire mirror.

mirror in a controlled way can the highest resolutions be achieved.

The mirror, which cost DM6 million, was made from a glass-ceramic blank at the West German firm Schott and polished at Carl Zeiss in Oberkochen, West Germany. At 24 cm in thickness, the blank is much thinner than a normal blank but still stable enough to provide decent images without active optics.

Active-optics technology is likely to play a significant role in the next generation of optical telescopes. Traditional, inflexible mirrors larger than 5 m in diameter tend to deform under their own weight, as astronomers in the Soviet Union have learned the hard way in attempting to build a 6-m telescope.

ESO's Very Large Telescope, an array of four 8-m telescopes to be built in Chile in the 1990s, will rely on active optics to manoeuvre mirrors even thinner than the NTT's. Researchers at Japan's Tokyo Astronomical Observatory plan a 7.6-m telescope using active optics to be constructed, probably in Hawaii, in the early 1990s.

Steven Dickman

Tunguska event celebrated

London

THE eightieth anniversary of the 'Tunguska event', the impact of a still-unidentified flying object in a remote region of eastern Siberia in summer 1980 has been marked by a scientific symposium at the Siberian Institute of Technology in Krasnoyarsk and by the inauguration of yet another expedition to the epicentre of the explosion. Although the symposium produced no major breakthroughs on the nature of the 'object', it did evoke a consensus that the time had come to abandon the exotic theories of the 1970s.

Ten years ago, acceptable hypotheses on the Tunguska object included a giant ball of ice and snow, antimatter, ball-lightning, a black hole, an alien spacecraft in distress and a nuclear explosion (possibly of extraterrestrial origin). This last suggestion came from one of the less conventional 'experts' — Aleksei Zolotov, originally an oil technologist co-opted by an academic expedition for his knowledge of the local terrain, who gradually emerged as an authority in his own right and who somehow never seems to encounter any problems in getting even his most bizarre theories published in Soviet journals. But even such an establishment figure as Professor Feliks Zigel of the Moscow Aviation Institute was prepared to argue, on the basis of the testimony of eye-witnesses, taken down long after the event, that the object had twice changed course in flight and hence was a spacecraft.

For the participants at this year's symposium, however, the time had come to abandon such fantasies. They called, instead, for a return to the fundamental data including a new survey of the site by both field expeditions and aerial photography teams. The last remnants of the debris from the explosion, they pointed out, are being swallowed up in the marshes, and there is no time to be lost. They urged that the proceedings of previous Tunguska symposia should be reprinted, and a catalogue of eyewitness testimonies drawn up.

The latest Tunguska expedition, which will work at the site for two months, will concentrate on a search for pulverized material which it is hoped can be identified as coming from the 'object'. (The absence of any meteorite-like body at the site has been a puzzle to investigators since the first expedition in 1927; hence the melting snowball and 'black hole' theories.) Further expeditions are planned for the next five years; laboratories in Moscow, Kiev and Novosibirsk will study the samples brought back, and the teams will report to the next Tunguska symposium in 1993.

Vera Rich

Signs of a respite for the bruised Australian science enterprise

Sydney

THE 1988–89 Australian budget is being hailed as a triumph for the treasurer, Paul Keating. In his sixth budget, announced last week, he has turned a deficit of A\$8,000 million in 1983–84 into a promised A\$5,500 million surplus. The hope is that with the budget back in balance, the worst will be over for science.

The treatment of the Commonwealth Scientific and Industrial Research Organization (CSIRO) is typical. With 7,200 employees, CSIRO is Australia's largest research body. This year it suffered its fifth successive year of cuts, losing A\$13.6 million from its A\$347.8 million budget of last year. But the government has promised that from now on the real value of support will be maintained and that future gains in outside revenue will not be taken away, giving CSIRO hope that there is light at the end of the tunnel.

Despite the cut, CSIRO's prospects should improve a little in the coming year. Completion of major capital works this year, such as the "Australia Telescope" radio telescope, will reduce the CSIRO's construction bill by A\$13.8 million next year, so that money for research will be held virtually constant in dollar terms; in real terms a predicted inflation rate of 5.5 per cent means a reduction.

Outside revenues for CSIRO from royalties, contract research, and sales of real estate are expected to increase by A\$27.8 million to around A\$129 million next year. This increase takes the total CSIRO budget to A\$463.2 million, up 6.5 per cent on last year and just keeping pace with inflation. The shift from government funding to industry funding necessarily entails a move from long-term basic research to short-term commercially oriented work.

The government says it wants CSIRO to find 30 per cent of its funding from outside sources by 1991. In the coming year the proportion should be 28 per cent. According to Keith Boardman, CSIRO's chief executive officer, the 30 per cent target might be exceeded. But he concedes that the morale of the organization, already very low, will suffer even more due to the inevitability of further staff reductions. Cuts over the last five years have resulted in a ratio of operating to salary expenditure which is unacceptably low.

In higher education, the budget gives effect to the reforms outlined by the Minister for Education, Employment, and Training, John Dawkins in a recent white paper (*Nature* 330, 592; 1987), including the unified national system and a graduate tax. Also in the budget was a plan to expand Australia's higher educa-

tion system by 10 per cent by 1991. The introduction of a graduate tax marks the end of free tertiary education in Australia. Under the "Higher Education Contribution Scheme" (HECS) each tertiary student will pay A\$1,800 for each year of their course. On average the annual charge will amount to 20% of actual course costs. Payment will be made as a levy of 1%, 2%, or 3% on income tax, according income, until the money owing has been repayed. No money need be repaid until the student has a job with an annual salary of A\$22,000 or more.

Research funds distributed annually by the Australian Research Council (ARC) are to be increased in steps, from A\$78 million in 1988 to A\$140 million in 1991.

Funding to the Australian Space Board is to be restored. The Space Board, established in 1986 with a budget of A\$5.3 million, had its funds slashed to A\$3.2 million in 1987. In 1988–89 it will receive A\$5.4 million.

Charles Morgan

Japan's mini-shuttle readied for flight

Tokyo

JAPAN'S Institute of Space and Astronautical Sciences plans to launch a space shuttle into the upper atmosphere later in the month from a balloon — but it is only a 2-metre model.

The model, developed by a team at the institute led by Professor Makoto Nagatomo, has already been test-flown in the lower atmosphere by releasing it from a helicopter (see *Nature* 322, 201; 1986). But the flight, scheduled for 18 September, will be the first attempt at re-entry into the atmosphere.

The shuttle will lift off from the institute's space centre in Kyushu cradled in a gondola and will be carried to a height of 20 km by a balloon. A booster rocket will then blast the vehicle to a height of 83 km where the booster will separate and the shuttle will splash down in the Pacific Ocean several minutes later.

The institute's ultimate aim is to develop a Highly Manoeuvrable Experimental Space (HIMES) vehicle equipped with a cryogenic engine. It will be about seven times the size of the model. HIMES could then be used as an unmanned orbital re-entry vehicle for scientific experiments.

But the HIMES project has yet to get the official backing of the education ministry. And the model shuttle is supported only with 'preparatory' research funds of a few million dollars a year — hence the balloon, a cost-saving measure. David Swinbanks

Mosquito release blocked by fearful California residents

Berkeley

A PUBLIC outcry from a Los Angeles, California, community has halted plans for an experiment to study population characteristics of mosquitoes in an urban environment. The experiment called for the release of some 20,000 mosquitoes in the city of Norwalk on 6 September. Ironically, an experiment involving the release of twice as many mosquitoes in a neighbouring community went off without a hitch earlier this summer.

Interest in Los Angeles mosquitoes, especially those in the Norwalk area, arises from an outbreak of St Louis encephalitis virus that occurred in 1984 and resulted in the death of a Norwalk resident. Richard P. Meyer of the University of California arbovirus field station in Bakersfield, California, says three mosquitoes found in the Los Angeles basin are known to act as vectors for the virus: *Culex quinquefasciatus*

(southern house mosquito), *C. tarsalis* (western encephalitis mosquito) and *C. peus* (foul water mosquito).

At the request of the Southeast Mosquito Abatement District, Meyer and his colleagues at the University of California School of Public Health in Berkeley have been studying these mosquitoes in an attempt to establish how the encephalitis outbreak occurred, and why it has not recurred. Meyer says one crucial factor is the extrinsic incubation period. The mosquitoes commonly pick up the virus from infected birds, but it takes some time before the virus appears in sufficient quantities in the mosquito saliva to be transmitted to another vertebrate. Warmer temperatures seem to shorten the incubation, and 1984 was a particularly hot, wet summer in Los Angeles.

Although there have been many controlled release studies in rural environments, urban mosquito studies are rare. For the Los Angeles basin experiments, mosquitoes were to be collected locally, tagged with a fluorescent dust, released from one or two locations and then recaptured in traps scattered around the city. That is how things worked in June for an experiment in Rossmore, up the San Gabriel river from Norwalk.

But not all urban environments are equally hospitable to mosquitoes. Unattended backyard pools, outdoor flower pots and drainage ditches make suitable breeding grounds, but these vary in number, in part because of the relative affluence of the community involved and how well its water supply is managed. More experiments were needed in a different kind of community.

The researchers chose Norwalk and, given their experience in Rossmore, expected no problems. But when Jack Hazelrigg of the Mosquito Abatement District appeared before the Norwalk City Council to explain the experiment, he was lampooned in the local press as planning to release "killer mosquitoes" on an unsuspecting populace with the blessing of local elected officials. Despite assurances that the experiment was extremely benign, and unlikely even to result in increased mosquito bites, let alone cases of encephalitis, the Mosquito Abatement District cancelled the planned release, and there are no plans to reschedule it.

Hazelrigg says that most people would probably never have noticed the experiment, and it is ludicrous to think that his agency would release "killer" mosquitoes. But a small, vocal minority has spoken up loudly, and, says Hazelrigg, "the squeaky wheel got the grease".

Joseph Palca

AIDS test for one-in-three newborns

Washington

THE United States has begun a programme to test nearly one-third of all newborn babies for the presence of antibodies to human immunodeficiency virus (HIV), the virus causing AIDS.

The study will provide information on the prevalence of HIV infection that is more representative of the general population than that being gathered elsewhere. Other studies are focusing on military recruits, patients at selected hospitals, women attending prenatal testing and abortion clinics and people seeking treatment at sexually transmitted disease clinics.

The tests will be administered as part of a standard series of blood tests to screen for metabolic disorders and infectious diseases that are performed shortly after birth. The results of the tests will be completely anonymous: the mother will not be told whether the results are positive or negative, and the data reported to the US Centers for Disease Control will contain only the mother's age and race. The zip code of the mother's address will be recorded so that regional assessments may be made.

Because there is no cure for AIDS, it is not considered unethical to perform anonymous blood tests for HIV without notifying the donor of the results. Anonymous testing is supported by both homosexual rights groups and the American Civil Liberties Union because it guarantees that the test results cannot be used to discriminate against individuals who test positive. Plans for a random seroprevalence survey of 45,000 US households which were announced at the Third International Conference on AIDS last year, and a 'pilot study' for the random survey based in Washington D.C., have been held up by protests over possible discrimination.

The newborn study is based on one begun in Massachusetts in 1986, which revealed that 2.1 in every 1,000 babies born alive in that state were infected with HIV. Thirty US states and the District of Columbia are expected to participate in the new study, so that a third of the babies born next year in the United States, estimated to number about 3.9 million, will be tested.

The state of New York also began testing newborn infants for antibodies to HIV last year (see *Nature* 328, 567; 1987). The New York study has so far shown a statewide infection rate of 8.6 in 1,000 and an infection rate for New York City of 16.4 in 1,000.

Carol Ezzell

Loma Linda halts controversial trial

Berkeley

LOMA Linda University Medical Center in southern California last month suspended its programme of keeping infants born only with a brain stem alive on life support systems in the hopes of using their organs for transplantation (see *Nature* 330, 592; 1987). Such anencephalic babies normally do not survive long after birth.

Joyce Peabody, chief of neonatology, said last week that twelve infants were studied under a protocol established late last year. At first, infants were placed on life-support systems immediately after birth, but after six births it became clear that life support intended to preserve the heart and lung was also supporting the brain stem, prolonging life. The final six infants were placed on life-support systems only when death was imminent, but this too interrupted the normal dying process, preserving some brain-stem function. Organs may be removed only following brain death, and no organs were obtained from the infants in the protocol that could be used for transplantation.

Loma Linda has had a vigorous programme of neonatal organ transplants. Peabody says that despite not obtaining usable organs, the programme was a success because it brought out some of the medical and social issues that will have to be considered in using anencephalic infants as organ donors. The medical centre has not decided whether it will resume the programme.

Joseph Palca

Social science defended

SIR—Your leading article "What happened to social science?" (*Nature* 334, 91; 1988) rightly draws attention to important gaps in the relationship between research and government policy-making. But it is hard to see what basis you can have for your belief that social scientists have "fallen into silence" over serious social issues. Without entering into the comparison you make with their former "vociferousness", it is certainly fair to say that they have been and are working on the issues of social relevance and public interest to which the latter half of your article refers.

Indeed, in each of the areas of social concern you identify, and of course in many others, the Economic and Social Research Council (ESRC) is funding research projects that seek to illuminate issues and provide analysis and recommendations for policy-makers, and a selection of these clearly shows how social scientists are exerting themselves. For example, the work of the Sir Norman Chester Centre for Football Research at the University of Leicester has analysed soccer hooliganism and works closely with the Department of the Environment and the football authorities to produce solutions. The ESRC has also commissioned studies on the problems of alcohol abuse: the Addiction Research Centre at the University of Hull has recently undertaken research for the Department of Health and Social Security on local prevention strategies.

The ESRC has also just started a major programme of research into the social and behavioural aspects of human immunodeficiency virus infection and AIDS, including studies of young people, the behaviour of drug users, the behaviour of bisexual men and the way AIDS is covered in the media. The aim of this research is to provide information on which effective programmes of health education, health policy and other intervention strategies can be based.

A further example is in the ESRC-funded research on the growing discipline of health economics, which is contributing to more effective and efficient health policies. Measures of the effects on health of various treatments, developed at the Centre for Health Economics at the University of York, are now being used within the health services to decide on priorities for courses of treatment.

These are just a few examples of the many important issues that social scientists are addressing with the help of funds from the ESRC. Certainly the council takes, and is addressing, the point that more needs to be done to ensure that social science findings are well disseminated and made available to busy policy-makers in a digestible form. For example, in the

autumn, the ESRC is holding a seminar involving policy-makers and academics to discuss how social science research can be better utilized in policy-making. We agree that the impact of research on policy is difficult to assess. There is an active social science research community concerned with policy, results and advice, and there are receptive audiences for their research. They need to be brought closer together and the ESRC is actively working for that improvement.

D.V. STAFFORD
(Acting Chairman)

*Economic & Social Research Council,
160 Great Portland Street,
London W1N 6BA, UK*

SIR—Social science is in disarray not because of government neglect (*Nature*, 334, 91; 1988) but because it has failed to deliver the goods. The promissory notes of the 1950s and 1960s that social scientists could solve the problems of crime, education, social prejudice, psychopathology and even warfare are unredeemed. Due to inappropriate models of human functioning, the problems have sometimes been exacerbated. The widespread belief that there is no genetic basis by which people differ in important ways so that discipline and punishment are counter-productive, has led to unbenign social policies, research programmes and editorial biases. It is rapidly becoming apparent from breakthroughs in human genome sequencing, behavioural genetics and developmental sociobiology that a very different state of affairs exists, one which has already challenged the way we think about life, and what is right and what is wrong. Until the social sciences 'come to terms' (as their practitioners like to say) with their biophobia and construct truly biosocial models of human functioning, they will continue, and perhaps ought to continue, to languish on the periphery.

J. PHILIPPE RUSHTON
*Department of Psychology,
University of Western Ontario,
London, Ontario N6A 5C2,
Canada*

Precinct NIH

SIR—The detailed analysis of the "Baltimore case" by John Maddox (*Nature* 333, 795; 1988) ended on an upbeat note, expressing confidence that the US Congress will no doubt leave it to the scientific community to take care of problems of misconduct in research. We should therefore note the following disquieting statement in the recent bill on National Institutes of Health (NIH) appropriations, approved by the Senate Appropriations Committee: "The committee

also expects NIH to strengthen its internal investigative responsibilities and fully explore the possibility of requiring the many peer review panels that evaluate research proposals to also on a random basis evaluate the results of that NIH-sponsored research with a focus on the broad area of scientific misconduct."

This suggestion seems too fantastic to get very far. Scientists volunteering to serve on review panels would surely be unlikely to welcome the assignment of becoming policemen, to consider it necessary, or to see how to accomplish the goal. Nevertheless as fraud is now being discussed as though it were a major problem in science, and as in the end the power of the purse is virtually unlimited, we cannot be sure. The proposal is certainly worth keeping an eye on.

BERNARD D. DAVIS
*Bacterial Physiology Unit,
Harvard Medical School,
25 Shattuck Street,
Boston,
Massachusetts 02115, USA*

Space defence

SIR—D.N. Spergel and G.B. Field (*Nature* 333, 813; 1988) give very persuasive arguments against the cost-effectiveness of space-based interceptor ballistic-missile defence. Such arguments may collapse, of course, if some unexpected new technological breakthrough should develop.

An argument which is less vulnerable to the possibility of such breakthroughs can be based on the fact that there are two possible outcomes of any such defence system with space-based components (SBCs):

(1) Locally concentrated anti-satellite (ASAT) forces could become capable of creating a window (in the necessarily dispersed SBCs) for attacking ballistic missiles to pass through.

(2) The SBCs could somehow be made survivable against ASAT attacks. But then the same space-survivability techniques would make possible at least equally survivable (and therefore unstoppable) nuclear orbit-to-Earth missiles (NOEMs), which would give much less warning time than ballistic missiles when launched against ground targets.

In either of these two scenarios, the defences would be overcome.

Early versions of NOEMs were tested by the Soviets in the 1960s. They are currently banned by the 1967 Outer Space treaty, but Gorbachev warned at the 1985 summit that all arms control "will be blown to the winds" if a ballistic-missile defence with SBCs (banned by the Anti-Ballistic Missile treaty) is deployed.

LOUIS A.P. BALÁZS
*Theoretical Physics Group, 50A-3115,
Lawrence Berkeley Laboratory,
Berkeley, California 94720, USA*

New ways with Bell's inequalities

Writing the rules that determine when quantum mechanics behaves non-locally in the language of information theory may carry important benefits, intuitive good sense particularly.

PEOPLE who design machines are applauded in their search for improvements, but those who design new theories are expected to get them right the first time. That, the Bohrs, Einsteins and even Newtons may have thought, is the burden natural scientists must carry. Certainly people pick over the foundations of successful theories long after they have become widely used. But the process of re-examination does not imply scepticism, and may even be a token of the respect a worthwhile theory commands.

That is the spirit in which the attention now, as for more than half a century, being paid to the foundations of quantum mechanics should be regarded. The process is also interesting. So much can be told, for example, from the continuing argument about the Einstein, Podolski & Rosen (EPR for short) *gedanken* experiment, first formally described more than half a century ago (*Phys. Rev.* **47**, 777; 1935) and intended as a demonstration *ad absurdum* that Bohr's doctrine of quantum measurement must be incorrect.

The principle is simple and general. Take some quantum system, say an electron and a positron, and suppose that it is transformed into two separately identifiable parts, say two photons, which become physically separated. There is nothing to prevent a measurement of the properties of one of the two entities (if a photon, its polarization). In Bohr's language, the value obtained by measurement can be any one of those accessible to the corresponding quantum system (plus or minus for polarization).

The EPR 'paradox' is then the rhetorical question, "How can the result of that measurement, arbitrarily one of an allowable set of values, be communicated to the other component of the system in such a way that the allowable results of a similar measurement on the second component are constrained so as to fit in with the overall dynamics of the system?"

The earlier quantum version of the Young's slits experiment raises the issue of non-locality in quantum mechanics. (Q. *Through which slit did the electron pass?* A. Through neither one nor the other, but through both.) The EPR paradox sharpens it. Einstein in particular argued for what the philosophers would call 'local realism' in physics, the notion that measurable properties of a physically separated part of a quantum system cannot seriously be determined by the measured properties of

another with which it once had a connection, but has no longer.

It is well-known that the argument continues. Perhaps the surprise is that it took until 1961 for J. S. Bell to put the issue succinctly enough for it to be tested experimentally. (There are experimental difficulties as well, as those who would prepare beams of polarized electrons or positrons know.) Bell's 'inequalities' are relations obtaining, for systems with several separately measurable components, between the correlations between measurements of the different parts. If local realism obtains, the inequalities are satisfied. With non-localization, they may be violated. There are now experiments to demonstrate violation, bringing aid and comfort to the Copenhagen school.

No doubt the 'Aspect experiment' ('Aspect' is a proper name, not an abstract noun), already six years old, would have settled the issue in favour of Copenhagen if it were not for the ingenuity required to devise quantum systems for which there is a realistic chance of demonstrating violation and for the acknowledged circumstance that Bell's inequalities are not a sharp test in the sense that only some departures from local realism will violate them.

That is why it is interesting and possibly important that Samuel L. Braunstein and Carlton M. Caves, from the California Institute of Technology and the University of Southern California, Los Angeles, have now devised a way of putting Bell's inequalities in the language of information theory and, in the process, of making them at least a little more general (*Phys. Rev. Lett.* **61**, 662; 1988). Formally, of course, there should be very little difference. The information content $I(a)$ of the measured value a of some quantity A is defined as $-\log p(a)$, where $p(a)$ is the probability of a and those who think in bits of information should use logarithms to the base 2. Because probabilities are multiplicative, information contents are additive; because probabilities are non-negative and less or equal to 1, information contents are non-negative, as the textbooks make plain. Because of the minus sign, this definition (due to Shannon) embodies the satisfying notion that measured values with low probability have high information content.

The trick is to use Bayes's theorem (the joint probability of a and b is the probability of a multiplied by the conditional

probability of b given a , or vice versa) repeatedly to construct relations between the average information content (averaged over all possible measured values) of measurements on two parts of a quantum system. Formally, it is like Bell's original device of calculating correlation coefficients.

One advantage is that the intuitive structure of probability arguments persists, for example in the information equivalent of Bayes's theorem that the average information content of two jointly measurable quantities A and B is the average information content of A plus the average information content of A given B . Another is that there are more joint information contents to play with — not merely the equivalents of joint and conditional probabilities, but the quantity called 'mutual' information which is the information embodied in common by two physical quantities. The calculations chime in neatly with the expectations of quantum mechanics in that, for example, the information content of B given A is always less than the information content of B itself — the measurement of a second component of a system is less revealing than that of the same measurement of the first component, which is what one would expect if quantum correlations are real.

What Braunstein and Caves claim for their technique, apart from analogues of the Bell inequalities, is that the calculations are more easily applied to complicated systems and that there may even be circumstances in which their inequalities are more immediately violated by departures from local realism. But they also argue explicitly that their formalism can be used to show that two counter-propagating particles with finite spin which are derived from a system with zero angular momentum must violate local realism for all values of the spin, although not necessarily in all conditions in which the measurements of the two spins might be carried out.

With all the effort devoted to the theory of quantum measurement since the mid-1920s, it is of course remarkable that these issues should only now be taken up. But it should not be surprising.

When it is only now, nearly three centuries after Newton, that people realize that classical mechanics admits of classical chaos, we may be lucky that so much attention is being paid to the basis of quantum mechanics. **John Maddox**

Protein structure

The shape of things to come?

Janet M. Thornton

THE prediction of the three-dimensional structure of a protein from its amino-acid sequence remains one of the fundamental unsolved problems in molecular biology. Because of the explosion in the number of known sequences compared with the rather lamentable number of structures, the need for accurate structure prediction is growing rapidly. This knowledge should help to elucidate how a protein works and should contribute to a more rational approach to drug and vaccine design and protein engineering. Since the first attempts at protein structure prediction, there has been a gradual shift in emphasis from prediction *ab initio*, using energy functions, to the more pragmatic approach of structure recognition by pattern matching. On page 45 of this issue¹, Rooman and Wodak explore rigorously by an automatic method the correlation between short sequence patterns and secondary structure. Rooman and Wodak have searched for sequence patterns that are consistently associated with a defined secondary structure, and then use the patterns for prediction. The good news is that short sequence patterns which reliably define secondary structure do exist. The bad news is that the prediction accuracy obtained using these patterns is still only about 60 per cent. The main problem is the lack of data to define an adequate number of patterns with strong predictive power.

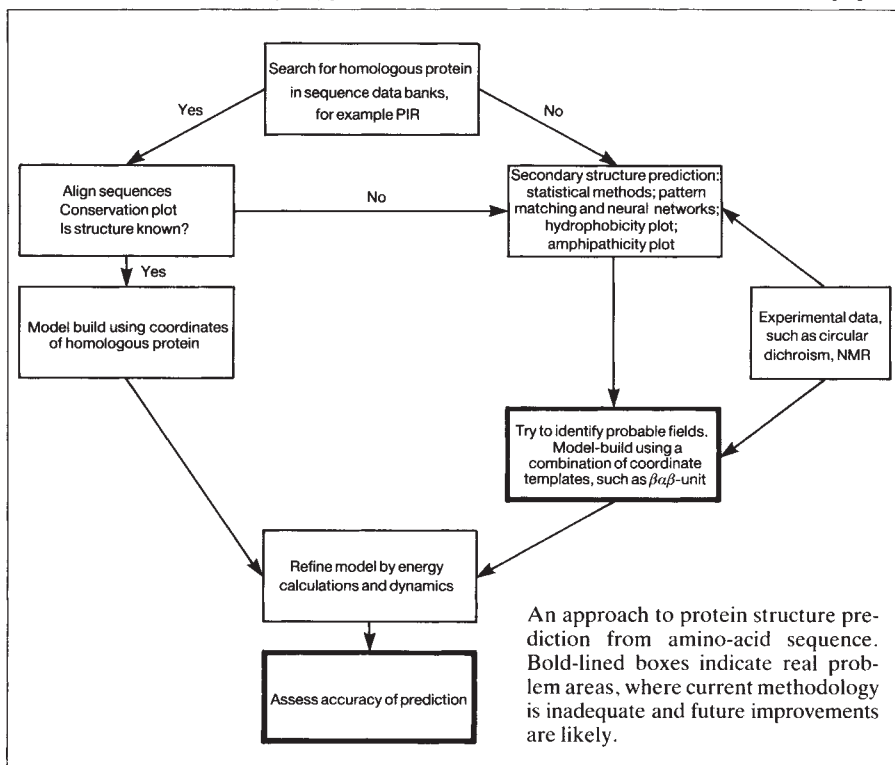
Initial attempts at protein structure prediction focused on evaluating and minimizing the energy of a protein using empirical energy functions. This approach failed because of inadequacies of potential functions, the 'egg-box' nature of the energy surface and lack of suitable minimization routines and inadequate computer power. The development of molecular dynamics and supercomputers have removed some of these problems, but even predicting the conformation of a relatively short loop with fixed end-to-end distances is still not straightforward.

Attention then shifted from these physically based methods to statistical analyses of known sequences and structures. The widely used secondary structure prediction algorithms (see refs 2,3) were first developed, using the less than 20 protein structures then available, based on singlet amino-acid properties. The accuracy of these methods has remained stubbornly fixed at about 60 per cent despite the increasing size of the database and using different methods in combination. Clearly, the singlet approach is fundamentally inadequate, because a

particular amino acid carries different information according to the residues in its environment.

Sequence patterns were first used to predict structure by Lim⁴, who derived patterns based on the packing of side chains in secondary structures. More recently, Cohen *et al.*⁵ have predicted coil regions using a complex set of patterns and some knowledge of protein architecture. They claim 95 per cent accuracy, which seems very promising. A more general pattern-matching approach to

of sequence patterns that are regularly associated with a specific secondary structure. The patterns studied are short (less than nine residues), with only two or three amino acids explicitly specified. Most patterns are only retained when 80 per cent of examples in the database have the same secondary structure assignments. Using 13,214 residues from 75 non-homologous proteins, they extract more than 22,000 sequence patterns. Only some of these patterns are strongly predictive. They find, for example, that Gly-Gly-X-Leu adopts the structure $\alpha\beta\beta$ in 12 of the 13 examples in the database (where X is any amino acid, α is any structure and β is a β -sheet). Another sort of pattern they distinguish is illustrated by the sequence Ala-Ala-X-X-Lys. In 14 of 16 examples in the database, this 5-residue peptide



secondary-structure prediction has been developed in several laboratories⁶⁻⁸ using the structures of homologous peptides in the database to predict the structure of an unknown sequence. This procedure again gives only about 60 per cent accuracy on average, either because some peptide sequences are not predictive, or because strongly homologous peptides do not occur in the current database. With the discovery that only 20 per cent of identical pentapeptides in unrelated proteins of known structure adopt the same secondary structure^{9,10}, the ability of short sequence patterns to predict structure needed investigation.

Rooman and Wodak in this issue¹ explore the question systematically. They have developed an elegant sequence and structure pattern specification allowing fully automated searching and extraction

includes 3 helical residues. Most patterns, however, are not so good and Rooman and Wodak find that in *de novo* prediction only 64 per cent accuracy is achieved at best. The strength of the new work lies in the careful analysis of which patterns contribute most to the prediction and why the success rate is so low.

The authors conclude that first, the reliability of prediction for a given pattern increases as the number of occurrences of that pattern in the learning set increases. Patterns which occur three times, for example, give correct predictions for only 40 per cent of hits. That is, some of the patterns are of high predictive power but others are not. In contrast, patterns which occur more than 60 times are correct in 85-90 per cent of predictions. Second, patterns involving the specification of three residues are inherently more

accurate than two residue patterns, as expected. Third, current database size (75 proteins used here) appears adequate for two-residue patterns for which predictive results have plateaued, whereas it is woefully inadequate for three-residue patterns. Most three-residue patterns occur less than three times in the database and so cannot be used predictively at all. It is suggested that patterns should occur about 15 times for accurate prediction.

The authors make some reasoned speculations that approximately 1,500 protein structures are needed for accurate prediction. Even with a generous estimate of 50 structures being elucidated per year, this process will take more than 20 years. It is therefore reasonable to ask if there is any chance of improving predictions within the next 10 years, by which time we will probably know the sequences of more than 100,000 proteins. The general approach to structure prediction is shown in the figure.

Several areas look promising. Techniques for aligning sequences are being extended to recognize very distantly related proteins and allow model building on the basis of a known structure. The use of consensus templates is likely to be particularly powerful (see ref. 11 for a review). At one extreme, this method may be able to identify similar folds in unrelated proteins. The accuracy of structure prediction can be improved by using a family of homologous sequences¹², and this technique should be used where possible.

Qian and Sejnowski in the current issue¹³ of the *Journal of Molecular Biology* show that neural network models, used previously for analysing speech patterns, can be applied to predicting secondary structure. The network model learns the relationship between sequence and structure from existing protein structures, and was then shown to give a 64 per cent accurate secondary structure prediction when applied to a test set of proteins. Qian and Sejnowski conclude that "no method based solely upon local information in the protein sequence is likely to produce significantly better results for non-homologous proteins".

In view of Roodman and Wodak's data, this conclusion seems unduly pessimistic. However, certainly the most obvious

limitation of most prediction algorithms currently available is that they do not yet adequately address the problem of tertiary structure prediction or try to incorporate any of the knowledge currently available about supersecondary and tertiary folds (see, for example, refs 5, 14). If we are to succeed before we have accumulated 1,500 protein structures, it will be necessary to encapsulate all our current knowledge including any available experimental clues. Structure prediction has remained elusive, but a problem of this importance deserves perseverance and

perhaps we would do well to heed the following advice (with apologies to Piet Hein¹⁵):

The road to success? why its plain and simple to express

Err and err

and err again

But less

and less

and less.

□

Janet M. Thornton is in the Department of Crystallography at Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK.

Seismology

The fossil roots of continents

Peter Shearer

THE oldest rocks exposed at the Earth's surface are found on the continental shields, the stable central regions of continents, where they have remained relatively undisturbed for billions of years. But the longevity and extent of deeper continental structures is currently controversial^{1,2}. Silver and Chan present seismic evidence on page 34 of this issue³ that a 200-km-thick layer beneath the Canadian Shield has been stable for over 2.5 billion years. If their hypothesis is correct, this layer survived the breakup of the supercontinent Pangaea 180 million years ago and subsequent rafting of the layer over the underlying mantle by continental drift. This supports ideas that continents have chemically distinct roots to at least 200 km deep, which are mechanically strong because of their low temperatures and possibly because they are strain-hardened by ancient deformation.

Silver and Chan³ examine the seismic phase 'SKS' recorded at nine stations in North America and Europe. This phase travels as a shear wave in the mantle, but as a compressional wave in the liquid outer core. Because of this path in the core, the polarization of the SKS wave at the receiver should always be parallel to the source-receiver azimuth. But Silver and Chan find that the arriving wave is often composed of two shear waves with different polarizations that arrive at slightly different times. This phenomenon, shear-wave splitting, is indicative of an anisotropic region within the Earth in which shear waves with different polarizations travel at different speeds. Shear-wave splitting has been observed in several studies⁴⁻⁷ but rarely with the clarity of Silver and Chan's observations.

By examining earthquakes at various places around the globe, Silver and Chan show that the anisotropy must be located near each recording site. The time difference between the two shear-wave arrivals at station RSON on the Canadian

Shield (Figs 1 and 2) is nearly 2 s and almost certainly arises from anisotropy in the upper mantle beneath the station. By analogy with measurements of crystal orientation in samples of upper-mantle material in ophiolites^{8,9}, Silver and Chan show that the most probable explanation for the anisotropy at station RSON is that the top 200 km of the mantle beneath the

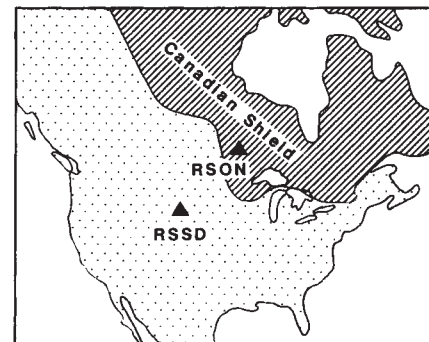


Fig. 1 Locations of the seismograph stations RSON on the Canadian Shield and RSSD in South Dakota.

station contains olivine crystals which are aligned at an azimuth of N 75° E.

Olivine crystals tend to become aligned in the flow direction during rock deformation¹⁰. The present motion direction of the North American plate at station RSON is not far from the observed anisotropy orientation. If the crystal alignment were related to the motion of the North American plate, the pattern of anisotropy should be relatively coherent over much of the plate, whereas Silver and Chan observe substantial variations in both magnitude and orientation of the anisotropy for stations only 1,000 km apart. They argue that a more probable explanation for the crystal alignment is the direction of the last significant deformation episode at each site. For station RSON, this occurred during the late Archaean (2.5 billion years ago) and the observed orientation of the Archaean geological fabric in surface rocks on the Canadian

1. Roodman, M.J. & Wodak, S.J. *Nature* **335**, 45-49 (1988).
2. Chou, P.Y. & Fasman, G.D. *Adv. Enzym.* **47**, 45-148 (1978).
3. Garnier, J. *et al. J. molec. Biol.* **120**, 97-120 (1978).
4. Lim, V.I. *J. molec. Biol.* **88**, 857-894 (1974).
5. Cohen, F.E. *et al. Biochemistry* **25**, 266-275 (1986).
6. Levin, J.M. *et al. FEBS Lett.* **205**, 303-308 (1986).
7. Nishikawa, K. & Ooi, T. *Biochem. biophys. Acta* **871**, 45-54 (1986).
8. Sweet, R.M. *Biopolymers* **25**, 1565-1577 (1977).
9. Kabsch, W. & Sander, C. *FEBS Lett.* **155**, 179-182 (1983).
10. Argos, P. *J. molec. Biol.* **197**, 331-348 (1987).
11. Taylor, W.R. *Prot. Engng.* (in the press).
12. Zvelebil, M.J. *et al. J. molec. Biol.* **195**, 957-961 (1987).
13. Qian, N. & Sejnowski, T.J. *J. molec. Biol.* **202**, 865-884 (1988).
14. Taylor, W.R. & Thornton, J.M. *J. molec. Biol.* **173**, 487-514 (1984).
15. Hein, P. *Brooks* (Borgens, Copenhagen, 1966).

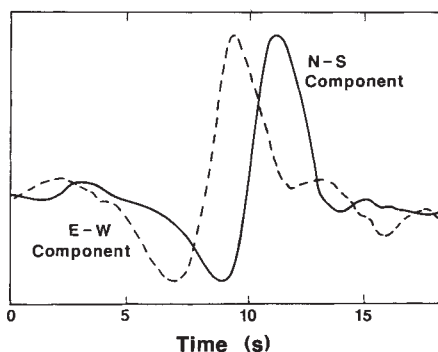


Fig. 2 Recordings from station RSON on the Canadian Shield show a 2-s delay between the arrival of the north-south and east-west components of the seismic phase 'SKS', caused by an anisotropic layer in the upper mantle beneath the station that splits the rising wave into two waves with different polarizations. In contrast, the splitting at station RSSD in South Dakota is only 0.6 s. (From Silver and Chan¹.)

Shield agrees with the anisotropy results. The observed orientation of the anisotropy at other stations also generally agrees with the grain direction for the last deformation event at each site.

The anisotropy observations at station RSON seem to show a crystal orientation which formed 2.5×10^9 years ago and has remained undeformed ever since. For this crystal fabric to have survived for so long, the mantle beneath RSON must be relatively cold, a result also suggested by observations of a high-velocity seismic root extending to a depth of 400 km beneath the Canadian¹¹ and Eurasian² shields. There seems to be a correlation between the thickness of the anisotropic layer and the presence of this high-velocity zone. At station RSSD in South Dakota (Fig. 1), where upper-mantle velocities are significantly less than at RSON¹¹, Silver and Chan calculate the anisotropic layer is only ~70 km thick.

These observations contradict models¹² in which the oceanic and continental temperature gradients are identical below 200 km. Instead, they support the theory that the low seismic velocities at depths of 200–400 km beneath the continental shields are related to temperature and compositional differences, and that this region moves coherently with the overlying plate^{2,13}. But other geophysicists¹ maintain that tem-

perature variations alone are responsible for the low velocities and that continental-plate motion is restricted to depths less than 200 km. (Details of these arguments were discussed by Thorne Lay in a recent News and Views article¹⁴.)

Regardless of the eventual outcome of this dispute, Silver and Chan's results provide strong evidence that up to 200 km of the Canadian Shield have remained relatively undisturbed for billions of years. This stability suggests that the region is mechanically very strong and Silver and Chan speculate that this strength may be a result of the Archaean deformation itself.

Conducting polymers

Seeing through synthetic metals

Martin R. Bryce

CHEAP, lightweight, durable and easy to make: these are the principal characteristics expected of polymer conductors that endear them to the electronics industry. Practice has not yet matched theory, but developments discussed at a recent meeting* suggest that there are grounds for hoping that polymers will replace metals as conductors in the next century. In particular, new, nearly transparent polymers have been synthesized that require less energy to excite electrons into the conducting state. Although these have lower conductivities than other polymers at present, they offer greater promise and are easily processed.

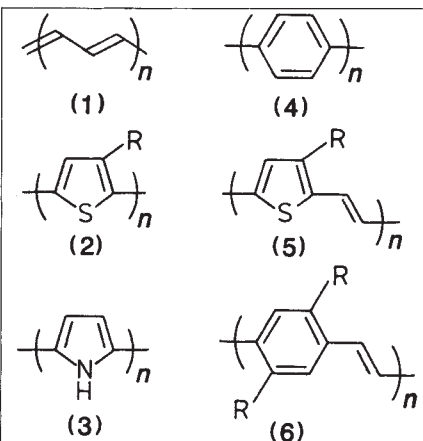
All organic polymers are intrinsic insulators, but high electrical conductivity in the range 1–500 siemens (S) cm^{-1} can be obtained when a conjugated organic polymer, that is one having alternating single and double bonds, is partially oxidized (p-doped) or reduced (n-doped) with suitable reagents — halogens and Lewis acids for oxidation; alkali metals for reduction. The usual polymers for these studies, polyacetylene (see 1 in the figure), polythiophene (2), polypyrrole (3) and poly-*para*-phenylene (4), are formed by electrochemical or chemical polymerization of the monomer units. By mechanically aligning the polymer chains of polyacetylene, conductivities as high as 10^5 S cm^{-1} have been achieved, comparable to that of copper (10^6 S cm^{-1}).

High conductivity and processability, both essential for industrial applications, were previously thought to be incompatible. But the addition of long, flexible chains — simple alkyl substituents ($\text{C}_6\text{--C}_{20}$ chains) or ether- and amide-containing chains (4–14 atoms long) — to the monomer units before polymerization renders polythiophenes soluble in common organic solvents, and hence proces-

In a process analogous to the cold-rolling of steel, the rocks beneath the Canadian Shield may have been strengthened by strain-hardening.

These shear-wave observations of fossil seismic anisotropy are based on seismograms recorded at just nine stations. The results are so promising that the technique will certainly become widely used and eventually seismologists may be able to produce detailed maps of the ancient deformation of continents. □

Peter Shearer is in the Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093, USA.



1, Polyacetylene; 2, polythiophene if R is an H atom (a) or R can be a long, flexible chain (b); 3, polypyrrole; 4, poly-*para*-phenylene; 5, polythiophene-vinylene (R is H, a, or an alkoxy chain, b); 6, poly-*para*-phenylene-vinylene (R is H, a, or $-\text{O}-\text{C}_6\text{H}_{13}$, b).

excitation energy (band gap) for the promotion of electrons into the conducting state, transparent or colourless materials should also be good conductors. Only one polymer is known which is both transparent and conducting, poly(isothionaphthalene), but this material is not processible.

R. Elsenbaumer (Allied-Signal Inc.) noted that polythiophene derivatives in which the heterocyclic rings are linked through a vinyl group (5a in the figure) have smaller band gaps than the parent polymer (2a) in which the rings are linked directly. Substitution of strongly electron-donating alkoxy groups onto thiophene rings also reduces the band gap. By

1. Anderson, D.L. *J. geophys. Res.* **92**, 13968–13980 (1987).
2. Lerner-Lam, A.L. & Jordan, T.H. *J. geophys. Res.* **92**, 14007–14026 (1987).
3. Silver, P.G. & Chan, W.W. *Nature* **335**, 34–39 (1988).
4. Booth, D.C. *et al. Geophys. J. R. astr. Soc.* **83**, 61–73 (1985).
5. Kaneshima, S., Ando, M. & Crampin, S. *Phys. Earth planet. Inter.* **45**, 45–48 (1987).
6. Ando, M., Ishikawa, Y. & Yamazaki, F. *J. geophys. Res.* **88**, 5850–5864 (1983).
7. Kind, R. *et al. Nature* **318**, 358–361 (1985).
8. Christensen, N.I. & Lundquist, S.M. *Geol. Soc. Am. Bull.* **93**, 279–288 (1982).
9. Christensen, N.I. *Geophys. J. R. astr. Soc.* **76**, 89 (1984).
10. Hess, H.H. *Nature* **203**, 629–631 (1964).
11. Grand, S.P. *J. geophys. Res.* **92**, 14065–14090 (1987).
12. Selater, J.G., Parsons, B. & Jaupart, C. *J. geophys. Res.* **86**, 11535–11552 (1981).
13. Jordan, T.H. *Rev. geophys. Space Phys.* **13**, 1–12 (1975).
14. Lay, T. *Nature* **333**, 209–210 (1988).

* International Conference on Science and Technology of Synthetic Metals, Santa Fe, 26 June–2 July, 1988.



100 years ago

IN a letter written on board the seal-ship *Jason* in the Denmark Sound, Dr. Nansen draws attention to the scarcity of seals on the coast of Greenland in recent years. Only ten years ago the animals were so plentiful and tame that thousands could be "clubbed" with the greatest ease. Dr Nansen is of opinion that the ruthless persecution of these animals since 1876, when the first sealer appeared in the Denmark Sound, has caused them to alter their habits. Formerly they were found on the edge of the drift ice, where they were safe from their only enemy, the Polar bear, though falling an easy prey to the sealer. Now they gather close to the shore, whither vessels cannot penetrate. From *Nature* 38, 422; 30 August 1888.

synthesizing a polythiophene (5b) containing both the linking vinyl group and the alkoxy substituent, Elsenbaumer confirmed his expectation that these two effects would be additive when present in the same polymer.

This polymer is deep blue in the insulating form, but on doping with electron acceptors (FeCl₃ and nitrosonium salts) its colour is reduced to a faint blue-grey; accordingly the band gap is about 1.32 electron volts (eV), comparable with that of *trans*-polyacetylene (1.3 eV) and about 0.4 eV less than the unsubstituted polythienylene vinylene PTV (5a). Moreover, the conductivity of the doped polymer, 2 S cm⁻¹, is still good. In the presence of suitable dopant anions, the alkoxy-vinyl derivatives (5b) are soluble in common organic solvents.

Similar vinyllogues (6a) and alkoxy-substituted vinyllogues (6b) of poly-*para*-phenylene (4) were reported by F. Wudl (University of California, Santa Barbara). The high-molecular-weight, stable prepolymer of the unsubstituted vinyllogue (6a) is soluble in organic solvents, giving processibility during the heat-induced polymerization reaction. Once formed, polymer (6a) is insoluble and cannot be processed. Successful processing at the prepolymer stage is unusual. It was first achieved for polyacetylene and facilitates the formation of well-crystallized, highly orientated films of conducting polymer.

By analogy with the thiophene derivative (5b), a soluble poly-*para*-phenylene derivative (6b) can be prepared if long chains are present. The hexyloxy chains are attached to the phenyl ring at the outset of the synthesis. Again the effects of vinyl linking and alkoxy substitution combine to reduce the band gap of the polymer. □

Martin R. Bryce is in the Department of Chemistry, University of Durham, South Road, Durham DH1 3LE, UK.

Natural selection

How females choose a mate

Ian Tomlinson

GAUDY male ornaments like those of the birds of paradise may have evolved because they are preferred by females in mating¹. There is now considerable evidence that the females of many animal species can choose their mates². Little is known, however, about how they express their preferences or what types of male character are chosen. Lynn Brodsky reports in a recent paper³ that females may prefer male characters that fit some general criterion rather than traits of a specific size, shape or colour.

Females cannot choose a male until

These mechanisms describe how a preferred mate is found, but they do not explain which males are chosen. There are four components of the process of mate choice that determine which male becomes the mate of a preferring female, and there are different mating behaviours possible within each component.

A female must have some innate, probably inheritable, standard by which she judges potential mates. This may instruct her to prefer a male of a particular value or type, or alternatively, one with as large a phenotypic value as possible. One

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Dressed to kill — male ruffs (Philomachus pugnax) displaying. (Photo K. Wothe/Bruce Coleman.)

they have found one or several potential mates. Perhaps individuals actively search for the type of male they find most attractive. Choice will tend to carry large costs, but preferred males may gain a large advantage because most females can express their preferences, with little dependence on male frequency.

Alternatively, a female may randomly encounter several males. Each encounter might, for example, raise her level of excitement towards some threshold that must be exceeded for mating to take place. If an encounter fails to push the female over that threshold, she will reject the male, perhaps again incurring considerable costs. An encounter with a preferred male might excite a female more than meeting a non-preferred male, and preferred males would thereby gain more mates⁴.

Another way for a female to find a mate is if several males display simultaneously to potential mates at a lek, for example. In species like the ruff (*Philomachus pugnax*; see figure), therefore, females can assess many males at once before making their choice. If preferring females rapidly decide on their mate, choice might carry few costs and could even be beneficial if non-preferrers take a long time to mate.

component of mate choice is how many males a female compares before she decides on her mate. At one extreme, she will accept the first male she finds that matches her standard: this is 'absolute' choice. At the other extreme, where preference is relative, a female will choose only after comparing very many males. She must adopt this strategy, for instance, if she wishes to select the male with the longest tail.

A second component is the strength of female preference, which may vary on a continuous or on an all-or-none scale. The preference of female tungara frogs (*Physalaemus pustulosus*) for large, loud-calling males seems to increase gradually with the strength of the call, for example⁵. Female two-spot ladybirds (*Adalia bipunctata*), on the other hand, seem to prefer melanic males independent of their degree of pigmentation⁶. A preference for a polygenic character might also be all-or-none, if it were expressed only above some threshold value of the male trait. A third component of mate choice determines the 'choosiness' of females — their keenness or ability to express a preference. Fitter females, for example, may be better able to be choosy.

Finally, preferences could differ in their

degree of 'fussiness'. A preference may be termed fussy when a male is chosen only if he fits some rigorous description. In the evolution of the peacock (*Pavo cristatus*), for example, females would prefer only males with large, 'eyed' tail coverts. Fussiness does not necessarily imply that a female has an all-or-none preference. A fussy preference for an ornament of a particular shape, size and colour might still vary continuously with the degree of development of that ornament. So, peacocks with large, eyed tail coverts would be preferred by females if they had more eyes.

Non-fussy preferences are unspecific and may simply result from perception biases. In her new work³, Brodsky provides evidence for non-fussy female preferences in the rock ptarmigan (*Lagopus mutus*). She shows that males with larger head combs are more successful at gaining mates, partly as a result of female choice. Comb colour varies slightly from scarlet to orange, but this does not affect mating success.

Brodsky randomly banded the legs of each male using different colours and observed mating success the following year. She finds that males given red or orange bands the previous year succeeded in gaining more mates than males with yellow, blue, green or black bands. She ascribed this to the choice by females of bands that resemble head combs in colour.

Three possible types of mating preference could explain these results. Female ptarmigan might simply have some liking for the colour red; or their perception may be biased towards red wavelengths; or preference genes 'for' red combs could instruct females just to look for some red male trait of a suitable size.

Brodsky's sample sizes are not large and her findings must consequently be treated with caution. It is also possible that red and orange bands act as a handicap to survival. In this case, males with red or orange bands that survive and return to mate the next year would be fitter than average, as they have been subjected to more stringent selection pressure than those banded with other colours.

Burley *et al.*⁷ have obtained similar results to Brodsky using the zebra finch *Poephila guttata*. They find that males wearing white hats or coloured leg bands are favoured by females. These authors suggest that females prefer band colours present in the male's natural markings, but it is not clear what factors have determined the evolution of the natural male coloration. Brodsky's work provides more convincing evidence for this idea by connecting a preference for an artificial male trait with a real male character that has probably evolved by female choice.

Brodsky's results are significant for the genetic control of mating preferences and

hence for the evolution of apparently maladaptive male ornaments. They emphasize that the term preference must be broadly defined. Female choice might not result from specific preference genes, but even if it does so, pleiotropic effects of those genes may be common.

A female preference for one male character might, therefore, also be a preference for many other traits that slightly resemble that character. Consequently, a very few preferences for general male characteristics could have caused the evolution of a set of complex, specific male ornaments. Perhaps the original preferred character was some advantageous trait like body size, possibly with low heritability. It could then be the case that the

complex and extreme male ornaments of animals like Count Raggi's bird of paradise, the widowbird, the ruff and the anolis lizards have evolved simply to fool females into thinking that a male is a larger, better-fed mate than he is in reality. □

1. Darwin, C.R. *The Descent of Man and Selection in Relation to Sex* (Murray, London, 1871).
2. Majerus, M.E.N. *Trends Ecol. Evol.* **1**, 1-7 (1986).
3. Brodsky, L.M. *Anim. Behav.* **36**, 662-667 (1988).
4. O'Donald, P. *Genetic Models of Sexual Selection* (Cambridge University Press, 1980).
5. Ryan, M.J. *Evolution* **37**, 261-272 (1983).
6. O'Donald, P. & Majerus, M.E.N. in *Mathematical Evolutionary Theory* (ed. Feldman, M.W.) (Princeton University Press, 1988).
7. Burley, N. *et al. Anim. Behav.* **30**, 444-455 (1982).

Ian Tomlinson is in the Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK.

Stellar evolution

What makes red giants?

Roger J. Tayler

STARS spend most of their lives as 'main-sequence' stars, converting hydrogen into helium in their interiors. But when their central hydrogen is exhausted, they become red giants in which hydrogen is converted into helium in a shell around a hydrogen-exhausted core. These luminous objects are 'red' because their surface temperature is such that their principal radiation is in the red part of the spectrum, and 'giants' because they are much larger than main-sequence stars of the same temperature. Astronomers have argued, without much agreement, about the principal physical reason for the existence of red giants. Applegate¹ now claims that stars become giants because their outer layers cannot transport the

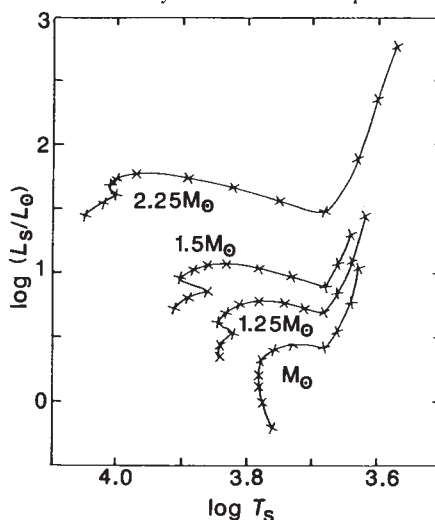
increasing luminosity produced in the hydrogen-burning shell, unless the envelope expands.

Early theories of stellar evolution showed that red giants must have strong radial gradients in their chemical composition. This idea has since been confirmed. But it was also thought that main-sequence stars remain chemically homogeneous as hydrogen is converted into helium, and accretion from the interstellar medium was invoked to explain the chemical gradients². It is now accepted that stars develop gradients naturally, the hydrogen in their cores eventually becoming exhausted.

At this point, the temperature in the core becomes uniform and, in the case of the more massive stars, when the core contains more than about 10 per cent of the stellar mass, it collapses under its own gravity³. But although this sequence accompanies the expansion of the outer envelope and the evolution to the red-giant phase, it is not clear that the composition gradient is the main factor which causes the expansion.

The contraction of the core produces an intense gravitational field near the centre of the star. Hoegpner and Weigert⁴ and Weiss⁵ suggest that the star's response to this field produces red giants. Weiss supports his argument by showing that a homogeneous star with a point mass inserted in its centre would be a giant.

According to Iben and Renzini⁶, the contraction of the core is not itself sufficient to produce a giant. But the way that the outer layers of the star respond to the increasing luminosity of the shell above the collapsed core can be unstable. When the instability occurs, the star's radius rapidly expands. The trigger for the instability is determined by the stellar



Post-main-sequence evolution of stars between 1 and 2.25 solar masses ($M_{\odot}$). The luminosity L_s is shown as a function of the surface temperature T_s . The bottom point on each curve is the main sequence. $L_{\odot}$ is the solar luminosity. The figure from ref. 9 is based on calculations by Iben.

opacity which governs the outward transport of energy by radiation and which depends on the star's temperature and density. If this is the true explanation of the giant phenomenon it depends on rather detailed physics. Iben and Renzini point out that it cannot explain all giants, for in stars less massive than about 1.2 solar masses, most of the energy in the relevant region is carried by convection.

Alternatively, the relationship between the pressure P and density ρ inside the stars could be a factor. This can be expressed in terms of an index n defined as a function of the pressure-density gradient: $n/(n+1) = d(\log \rho)/d(\log P)$. Although the interdependence of ρ and P , and hence n , varies inside a star, some insight can be gained by studying models with constant n .

In homogeneous stars, n is usually in the range 1.5–3; as $n \rightarrow 5$, the model radius of the star tends to infinity. Eggleton and Faulkner argued that the combination of a shell of hydrogen burning and a discontinuity in chemical composition at the core surface can produce a local value of n greater than 5, which is sufficient to produce a very large star⁷. Or, according to Yahil and van der Horn, a steep density gradient, equivalent to a large n , at the edge of the core generates the same effect⁸.

Applegate¹ now argues that it is the increase in luminosity of the hydrogen-burning shell that is the crucial factor. There is a maximum luminosity which can be carried by an overlying envelope whose opacity is given by Kramers' formula. If this luminosity is exceeded, the star expands rapidly until the energy can be radiated away. His detailed discussion applies to a star the mass of the Sun, for which the arguments of Iben and Renzini⁶ are not valid. Although Applegate singles out the rise in luminosity as the crucial factor he indicates that there must be an abrupt reduction of density at the surface of the core, in agreement with the n -dependent models^{7,8}.

I suspect that the argument about what causes what is not yet over. As there should be a few thousand million years before the Sun becomes a red giant and extinguishes life on Earth, there will be time for some further thoughts. □

1. Applegate, J.H. *Astrophys. J.* **329**, 803–807 (1988).

2. Hoyle, F. & Lyttleton, R.A. *Mon. Not. R. Astr. Soc.* **102**, 218–225 (1942).

3. Schönberg, M. & Chandrasekhar, S. *Astrophys. J.* **96**, 161–172 (1942).

4. Hoepfner, W. & Weigert, A. *Astr. Astrophys.* **25**, 99–103 (1973).

5. Weiss, A. *Astr. Astrophys.* **127**, 411–412 (1983).

6. Iben, I. Jr & Renzini, A. *Phys. Rep.* **105**, 332–406 (1984).

7. Eggleton, P.P. & Faulkner, J. in *Physical Processes in Red Giants* (eds Iben, I. Jr & Renzini, A.) 179–182 (Reidel, Dordrecht, 1981).

8. Yahil, A. & Van der Horn, L. *Astrophys. J.* **296**, 554–564 (1985).

9. Tayler, R.J. *The Stars: Their Structure and Evolution* (Wykeham, London, 1970).

Roger J. Tayler is in the Astronomy Centre, University of Sussex, Brighton BN1 9QH, UK.

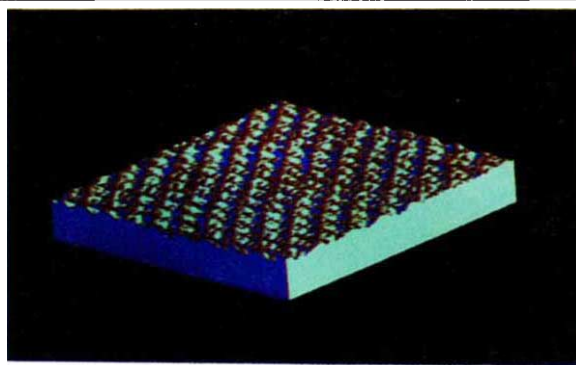
Microscopy

Electrons that make waves

C.F. Quate

SCANNING tunnelling microscopes, recently developed high-resolution devices that trace the form of surfaces using an atomic-scale tip, typically reveal atomic and molecular details, often with surprising clarity. But because the technique is sensitive to variations in the density of charge at the sample surface, usually determined by the ion positions, other electronic effects can also become

maximum electron density surround the nuclei of the atoms or are associated with molecular orbitals involved in the bonding between atoms. For these surfaces, the observed topography directly reflects the atomic structure. For other surfaces the periodicities in distribution of charge can differ from the periodicity of the lattice, and for these surfaces topographical data are not available. The work of Wu *et al.*



Projected view of an image of TaS₂ taken with a scanning tunnelling microscope at 236 K. The large hill-like structures are the maxima of the triclinic charge density wave (CDW) present at this temperature and the smaller modulation arises from the surface sulphur atoms (lattice constant 0.33 nm). In this view, the discommensurations forming the boundaries between two commensurate CDW domains are clearly seen as jogs introduced into the lines of CDW maxima. (For the best view of the discommensurations, hold the page up to eye level and look along the diagonal.) The image is 15 nm square. (Courtesy of John Clarke.)

apparent. On page 55 of this issue¹, Wu, Zhou and Lieber present images of charge density waves, spontaneous fixed or mobile corrugations of the charge distribution, at the surface of the metal tantalum disulphide.

In the scanning tunnelling microscope (STM), a sharp metal tip is placed close to the sample and mechanically scanned across the surface. The tip is tapered to a single atom at the apex, which sets the limit on resolution. The apex is held 5–10 Å from the surface, a distance comparable with the spacing between atoms in the solid (2–3 Å). In this configuration, the moving electrons can tunnel either to adjacent atoms in the surface or to the single atom at the tip. The current to the tip is proportional to the density of electronic charge at the characteristic 'Fermi' energy of the conducting electrons. The charge density can be imaged by scanning the sharp tip across the surface of the sample. Contours of constant brightness in the image correspond to contours of constant electron density on the surface.

On the surface of semiconductors and of many other materials, the contours of

reported in this issue¹ is a graphic illustration of the distinction between these two situations. The authors studied the surface of tantalum disulphide (TaS₂), a material easily cleaved to expose a fresh, clean surface that is inert and not easily oxidized. At high temperature, TaS₂ is a metal with a resistivity that is 100 times that of copper. At lower temperatures, the material undergoes a first-order phase transition to a new electronic state^{2,3}. The new phase, called a charge density wave (CDW), is a result of the strong coupling between the electrons and the vibrating ions in the lattice. The distribution of charge in the new state is modulated with a periodicity much larger than the periodicity of the ions. Superimposed on the ionic lattice there is an 'electron lattice' that is revealed in the new STM images of Wu *et al.*. The ionic lattice, determined by the positions of the atoms in the solid, does not show in the images.

In a second experiment, Wu *et al.* use a crystal chemically modified by replacing 10 per cent of the tantalum atoms with titanium atoms. With this substitution, the CDWs are suppressed and the ionic lattice reappears in the STM images, furnishing dramatic evidence that the tunnelling current in the STM is a measure of the surface charge density.

Electronic phase transitions, such as the transitions to ferromagnetism, to superconductivity and to charge density waves, are problems central to solid-state physics. The transition to CDWs is a significant effect: nearly one electron from each atom is transferred to the new state¹ producing the strong modulation observed in the charge density.

When first introduced¹, charge density waves seemed to be very esoteric objects

with little relevance to the real world. They can be seen in X-ray diffraction patterns as satellite 'wings' to the principal Bragg diffraction peaks generated by the ion lattice. They can dramatically alter the resistivity and magnetoresistance (variations of resistance in applied magnetic fields) of aligned metal wires. And because CDWs are a result of interactions between electrons and ion-lattice vibrations, they can affect the elasticity of samples.

Coleman, Hansma and co-workers were the first to show⁵ that all forms of CDWs can be studied with the STM. Several groups⁶⁻⁸ find that the properties inferred from diffraction studies⁴ can be obtained directly from the STM images. It turns out also that a small defect in the substrate can suppress the CDW near the defect⁹.

Clarke and co-workers⁹ have studied the four different charge density wave phases that occur in TaS₂ in the range of temperatures 77–370 K. At the highest temperatures the waves are incommensurate — the wavelength is not an integral multiple of the ion-lattice spacing. At the lowest temperatures, the waves are wholly commensurate. In between there are two nearly commensurate phases, one of

which (the triclinic phase, see figure) is obtained on warming only. Significantly, the transition temperatures at the surface as measured with the STM coincide with those measured for the bulk, meaning that CDWs extend to the surface layer unaltered in character.

This body of work is extensive but incomplete. The motion of CDWs has not yet been observed with the STM. In materials such as NbSe₃, a modest electric field will dislodge the standing waves of charge density from their 'pinning' sites and convert them to travelling waves that carry current through the sample. It would be fascinating to study this mobile "lattice within a lattice"¹⁰ with the STM. □

1. Wu, X.-L., Zhou, P. & Lieber, C.M. *Nature* **335**, 55–57 (1988).
2. Wilson, J.A., DiSalvo, F.J. & Mahajan, S. *Adv. Phys.* **24**, 117–201 (1975).
3. Comès, R., Lambert, M., Launois, H. & Zeller, H.R. *Phys. Rev.* **B8**, 571–575 (1973).
4. DiSalvo, F.J. & Rice, T.M. *Phys. Today* **32**, 32–38 (1979).
5. Slough, C.G. *et al. Phys. Rev.* **B34**, 994–1005 (1986).
6. Tanaka, M. *et al. Proc. R. micr. Soc.* (in the press).
7. Gammie, G. *et al. Proc. R. micr. Soc.* (in the press).
8. Coleman, R.V. *et al. J. Vac. Sci. Technol.* **A6**, 338–343 (1988).
9. Thomson, R.E. *et al. Proc. R. micr. Soc.* (in the press).

C.F. Quate is in the W.W. Hansen Laboratories of Physics, Stanford University, Stanford, California 94305, and in Xerox PARC, 3333 Coyote Hill Road, Palo Alto, California 94304, USA.

RNA processing

The post-RNA world

Miranda Robertson

INTRONS, in the light of the recently recognized catalytic properties of RNA, have come to be widely regarded as the legacy of a primordial RNA world¹, now largely expunged by bacteria in the interests of economy, exploited by metazoa for versatility, and tolerated by yeast. As part of this general scheme, it has been argued² that the self-splicing introns of latter-day organisms such as *Tetrahymena*³ are the direct descendants of autocatalytic introns in the original RNA genome; and that in the course of evolutionary time the catalytic function of introns has escaped into the ribonucleoprotein particles (snRNPs) that in modern metazoa mediate their removal, leaving behind only a minimum of conserved bases to guide cleavage and splicing. Yeast introns, whose sequences are more conserved, can be interpreted as an intermediate-to-late stage in the release of the intron from the constraints imposed by its catalytic origins; and other variant species of intron (see Fig. 1) can be understood as intermediates of different kinds in the same progression².

The quest for unity in the diverse taxonomy of surviving introns is of more than academic interest; for if yeast and mammalian mechanisms (for example)

are only minor variants of one another, then the mammalian mechanism becomes susceptible to investigation by yeast genetics, and the extensive biochemistry of mammalian splicing pathway components becomes available to yeast geneticists.

On the whole, the unified view has been increasingly upheld by the search for homologues of the mammalian snRNPs, a trend that was maintained at this year's Cold Spring Harbor Meeting on RNA processing^{*} where the differences between mammals and yeast continued to evaporate while trans-splicing, an eccentricity of trypanosomes and nematodes, emerged as an interesting variation on the metazoan theme, as foreseen by Sharp⁴; and an important focus for the future was signalled by the inclusion of a talk on the central part played by alternative splicing in the determination of sex in *Drosophila*.

Taxonomy of introns

Introns are removed from mammalian RNA precursors (pre-mRNAs) by two successive cleavage and splicing reactions. In the first, a conserved guanosine at the

5' end of the intron is joined to an adenosine a short distance from its 3' end to form a lariat, freeing the 3' end of the upstream exon. In the second reaction, the 3' end of the upstream exon is joined to the 5' end of the downstream exon with the release of the intron lariat (see Fig. 1). These reactions are mediated by snRNPs designated U1, U2, U4, U5 and U6 whose assembly into a spliceosome is outlined in Fig. 2 and its legend.

Yeast introns are distinguished from mammalian introns principally by the possession of a conserved sequence termed the UACUAAC box which specifies the branch point for the lariat and to which U2 binds by complementary base-pairing⁵. Although mammalian introns do not contain UACUAAC boxes, a degenerate consensus sequence can be derived from bases adjacent to the adenosine at the branch point; and that and the fact that mammalian cells will remove yeast introns has strongly suggested that despite the difference in sequence conservation, and the fact that the yeast U2 is six times larger than its mammalian homologue, branch formation in the two is essentially the same process.

Further support for this inference has been provided by manipulation of the branch-point sequences in synthetic globin constructs in which the first exon is joined to tandem duplicates of intron 1 and exon 2 so that one 5' splice site has a choice between tandemly duplicated 3' sites. Normally, exon 1 will become spliced to the closer of the two 3' sites; but Robin Reed (Harvard University) reported that alterations to the adjacent branch-point sequence (presumably disrupting complementarity to U2) can induce splicing to the distal exon. Using a similar construct, Yuan Zhuang (Yale University) has been able to achieve the same result by inserting a UACUAAC box in an appropriate position in the more distal intron. Thus, mammalian U2 has the same sequence preference as the yeast U2 but is more tolerant. Moreover much of the substantial size difference between the yeast and the mammalian U2 RNA is attributable to a ~ 1-kb dispensable domain in the yeast U2 that intervenes between two regions of conserved secondary structure (Elizabeth Shuster, University of California, San Francisco), and so seems unlikely to signify any fundamental functional distinction⁶.

Plethora of proteins

Plainly, complementary base-pairing plays an important part in the binding of U2 (and other snRNPs) to intron sequences; but an increasing number of protein factors is now known to have intervened in the assembly of the spliceosome since its (putative) ribozymal origin. One such, U2 snRNP auxiliary factor (U2AF), is thought to account for a second import-

*The Cold Spring Harbor Meeting on RNA Processing, Cold Spring Harbor, 11–15 May 1988.

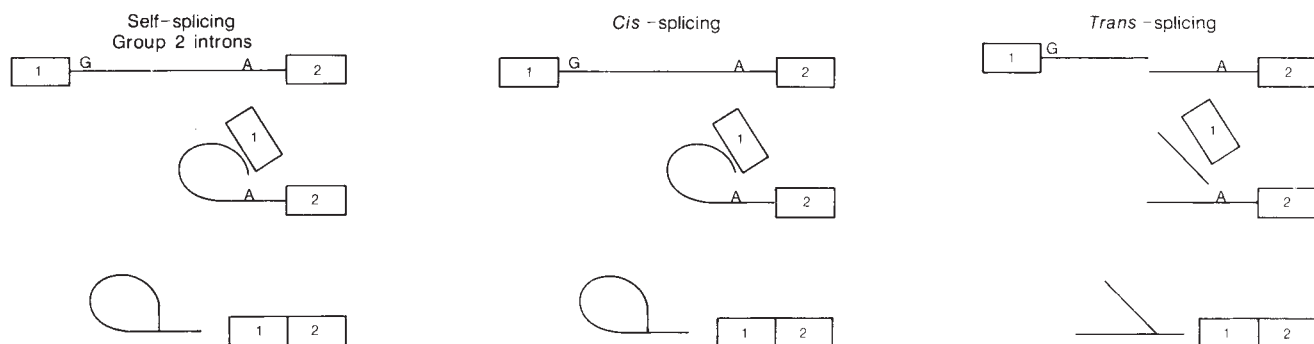


Fig. 1 Skeletons of the splicing reactions in self-splicing group 2 introns, with the *cis*-spliced higher eukaryotic and *trans*-spliced trypanosome and nematode configurations shown to illustrate their general similarities. In all three cases, the reaction depends upon nucleophilic attack by a conserved guanine nucleotide at the 5' splice junction on an adenine close to the 3' junction, with the formation of a branched intermediate. Group 1 self-splicing introns, such as those of *Tetrahymena*¹, undergo a somewhat different reaction that, for reasons no-one has yet explained, does not seem to have survived in higher eukaryotes.

ant difference between yeast and mammalian splicing requirements⁷. Whereas in yeast the efficiency of splicing depends primarily on the conservation of the UACUAAC box, in mammalian cells it depends critically on the distance of the branch point from the 3' splice junction, which is required for spliceosome formation in mammalian but not in yeast systems. U2AF binds to the 3' splice site and is required for U2 binding to mammalian introns, and may thus be responsible for the dependence of mammalian splicing systems on the distance between the branch point and the 3' splice site.

Other protein factors are now emerging from the fractionation of nuclear extracts in such numbers as to generate complaints that the spliceosome is coming increasingly to resemble the ribosome not only in size and composition but also in complexity and intractability. Yeast genetics may help with the definition of the fundamental protein components of the splicing pathway; and *Drosophila* genetics with the elucidation of differential splicing.

Many participants at the meeting took courage for example from the convergence of research by Jean Beggs (Edinburgh University) on the product of RNA8, a gene originally identified by Hartwell as one of a series implicated in RNA metabolism, and by Phil Sharp (Massachusetts Institute of Technology) on protein factors that bind at specific stages of spliceosome assembly. RNA8, according to Beggs, encodes a protein of relative molecular mass 260,000 (260K) that co-precipitates with U5 and is associated with the assembled spliceosome (see Fig. 2), as well as with the released intron. She finds that antibodies against the yeast protein recognize a protein of about the same molecular weight in HeLa cells.

At the same time, Mariano Garcia-Blanco from Sharp's laboratory reported that one of two proteins they have recently identified by ultraviolet crosslinking on pre-mRNA *in vitro* has a molecular weight of more than 200K, and which is implicated in the later stages of spliceosome

assembly and can be precipitated by antibodies against the RNA8 product. The same protein has been detected by Reinhard Lührmann (Max-Planck Institute, Berlin) in association with extensively purified U5.

It may turn out that all the essential invariant protein factors in higher eukaryotic spliceosomes have homologues encoded by yeast RNA genes, of which several more were reported by U. Vijayraghavan from Abelson's laboratory (California Institute of Technology).

U2AF may not be an exception because of the absence of distance sensitivity and the requirement for a 3' splice site in the formation of yeast spliceosomes. Ruskin *et al.* have argued⁷ that U2AF is absent from yeast; and if so, it may be implicated in the mechanism that allows differential splicing in higher eukaryotes.

Sex by splicing

It is not unreasonable to suppose that the diversity and complexity of multicellular life forms arose through the flexibility conferred by the evolution of differential exon usage. Its importance in the production of the specialized proteins characterizing the differentiated states of vertebrate cells is well established; and in *Drosophila*, it is clear that much of the posterior body pattern is determined by developmentally regulated differential splicing of the *Ubx* gene. None of the known mutations of *Ubx*, however, specifically affects differential RNA splicing.

There is, on the other hand, clear evidence that mutations affecting sex determination in *Drosophila* act directly on a hierarchy of sex-specific differential splicing reactions⁸⁻¹⁰. Sex in *Drosophila* is genetically determined by the ratio of X chromosomes to autosomes. The X:autosome ratio is 'read' by the product of the *sex lethal* (*sxl*) gene, whose activity is essential in females but completely unnecessary in males. It has been inferred from genetic evidence¹¹ that *sxl* positively regulates its own production as well as the activation of a second gene, *transformer*

(*tra*), that like *sxl* is necessary for female but not for male development. The *transformer* gene in turn controls *doublesex* (*dsx*), a gene that is expressed both in females and in males but in different forms, and whose products suppress the expression of male somatic characters in females and of female characters in males (see Fig. 3).

An analysis of the transcripts made in males and females from *sxl*, *tra* and *dsx* has shown that all three genes are transcribed in both sexes, but that the transcripts are differentially spliced. In the case of *dsx* this (presumably) gives rise to a functional transcript in both sexes¹²; in the case of *sxl* (as reported by L. Bell, Princeton University) and *tra*¹⁰, it turns out to be a device for inactivating the genes in males by the introduction of an

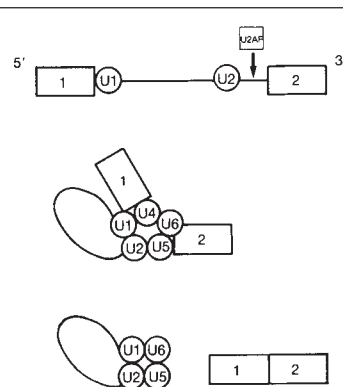


Fig. 2 The steps in the assembly of the spliceosome during cleavage and splicing of a higher eukaryotic intron. U1 binds at the 5' junction and U2 at a consensus sequence containing an adenosine residue about 30 nucleotides from the 3' junction where U2AF binds. In the second step, U1 and U2 come together with U4, U5 and U6 and a branched intermediate with the form of a lariat is generated by the splicing of the 5' end of the intron to the conserved adenosine, with the release of the 5' exon. In the third step, the 5' exon is joined to the 3' exon with the release of the lariat and the dissociation of U4. The lariat is then degraded.

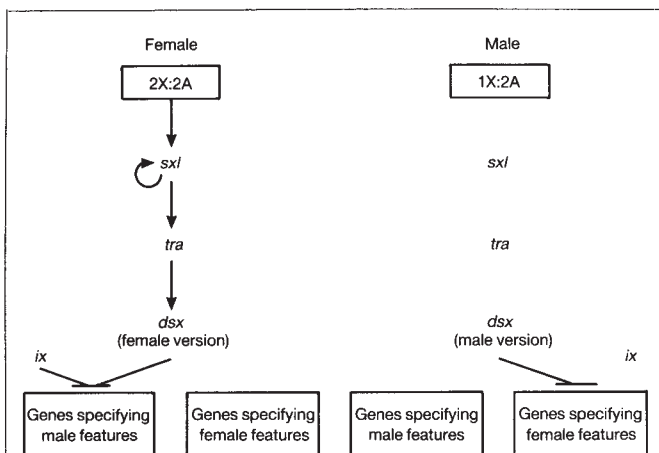


Fig. 3 Sex determination in *Drosophila*. In fruitflies with two X chromosomes, *sxl* is activated and activates itself and *tra* which generate the female version of *dsx* which, with *intersex* (*ix*), suppresses the expression of those genes that determine the male characteristics. In fruitflies with one X chromosome, *sxl* is inactive, *tra* is not activated and the male form of *dsx* is produced by default. The male variant of *dsx* suppresses the genes determining female characteristics.

exon containing a stop codon. Because mutations in *sxl* or *tra* cause aberrant patterns of splicing in *tra* and *dsx*, they are presumed to encode components of the RNA processing machinery critical for determining exon usage; and the characterization of their products may converge with the output from mammalian biochemistry in much the way that seems already to be happening in yeast.

Evolutionary intermediates

It is tempting to reflect that the evolution of sex seems easier to understand if divergence from a (purely hypothetical) hermaphrodite state could be achieved by the combination of an adventitious stop codon with some modification to a component of the splicing apparatus. The flexibility argument for retaining introns moreover gets a further filip from trypanosomes and from nematodes, both of which seem to have evolved the same strategy for economizing on exons by using the same one to encode the 5' leader sequence of a number of different genes through splicing the 5' transcript to the rest of the pre-mRNA in *trans*^{13,14}.

Trans-splicing of artificial constructs in a mammalian system *in vitro* is very inefficient in the absence of any significant complementarity between the introns undergoing the reaction (see ref. 4). There is no such sequence complementarity between the intron portions of the spliced leader transcripts and those of the RNAs to which they become joined; instead, each 5' leader sequence seems to function as an snRNA.

James Bruzik in Joan Steitz's laboratory at Yale University reported that in a comparison of six species of trypanosomes (or related protozoans) and the nematode *Caenorhabditis elegans*, similar secondary structures can be derived for all the spliced leader transcripts despite the absence of

share with snRNPs not only their characteristic relationship with the Sm antigen^{16,17} but also their trimethylguanine cap. The cap is inappropriate however for the mature RNA; and indeed Thomas and colleagues have shown that it is removed or modified in the course of maturation¹⁷. This, they argue, is perhaps the strongest existing evidence for the snRNA character of the leader since the additional processing step is very difficult to explain unless the trimethylguanine cap is important for *trans*-splicing.

This and some direct evidence from David Hirsh's laboratory¹⁸ for the existence of branched intermediates of the form shown in column 3 of Fig. 1 startlingly vindicates Sharp's speculation that the solution to the efficiency problem in *trans*-splicing might be just such a built-in snRNP; and makes the idea of snRNAs as the escaped descendants of self-splicing group 2 introns virtually irresistible. □

1. Gilbert, W. *Cold Spring Harbor Symp. quant. Biol.* **52**, 901-905 (1987).
2. Sharp, P. *Cell* **42**, 397-400 (1985).
3. Zaug, A.J., Grabowski, P.J. & Cech, T.R. *Nature* **301**, 578-583 (1983).
4. Sharp, P. *Cell* **50**, 147-148 (1987).
5. Parker, R., Siliciano, P.G. & Guthrie, C. *Cell* **49**, 229-239 (1987).
6. Igel, A.H. & Ares, M. Jr *Nature* **334**, 450-453 (1988).
7. Ruskin, B., Zamore, P.D. & Green, M.R. *Cell* **52**, 207-219 (1988).
8. McKeown, M., Belote, J.M. & Baker, B.S. *Cell* **48**, 489-499 (1987).
9. Boggs, R.T., Gregor, P., Idriss, S., Belote, J.M. & McKeown, M. *Cell* **50**, 739-747 (1987).
10. Nagoshi, R.N., McKeown, M., Burtis, K.B., Belote, B.M. & Baker, B.S. *Cell* **50**, 739-749 (1987).
11. Maine, E.M., Salz, H.K., Schedl, P. & Cline, T.W. *Cold Spring Harbor Symp. quant. Biol.* **50**, 595-604 (1985).
12. Baker, B.S. & Wolfner, M.F. *Genes Dev.* **2**, 477-489 (1988).
13. Van der Ploeg, L.H.T. et al. *Nucleic Acids Res.* **10**, 593-609 (1982).
14. Krause, M. & Hirsh, D. *Cell* **49**, 753-761 (1987).
15. Bruzik, J. P., Van Doren, K., Hirsh, D. & Steitz, J.A. *Nature* (in the press).
16. Van Doren, K. & Hirsh, D. *Nature* (in the press).
17. Thomas, J., Conrad, R. C. & Blumenthal, T. *Cell* **54**, 533-539 (1988).
18. Bektesh, S. & Hirsh, D. *Nucleic Acids Res.* (in the press).

Miranda Robertson is Biology Editor of Nature.

Daedalus

Franklin flies again

ONE of the annoying features of temperate weather, especially during a drought, is that most clouds fail to deposit rain. Many experiments with cloud seeding have been tried; Daedalus now proposes a new one. He recalls the classic (and highly dangerous) experiment conducted by Benjamin Franklin: flying a kite in a thunderstorm to conduct atmospheric electricity down the kite-string. Daedalus wants to try it in reverse. By pumping electric charge up the string, he hopes to trigger a thunderstorm.

A moist atmosphere is often potentially unstable. Once a parcel of warm moist air begins to rise, its water vapour condenses to cloud droplets; the resulting release of latent heat lifts the air further still, drawing in more moist air from the same stratum and triggering a general convective turnover. If the condensing cloud droplets could coalesce they would fall as rain.

So DREADCO's meteorologists are flying a captive balloon on a copper cable among suitable summery clouds. The balloon is equipped with many sharp corona-discharge points, and can be charged up from a big Wimshurst machine on the ground. By cranking the Wimshurst machine at the right moment, a sizeable electric charge can be injected as corona ions into a chosen cloud. Several things can then happen. The ions may nucleate droplet condensation by the cloud-chamber effect, thickening the cloud and releasing latent heat of condensation. If the cloud is below freezing-point, ice crystals may be nucleated instead: these rapidly condense water from a cloud and aggregate it into rain. Even without nucleation, charge injected into a parcel of air will expand it by self-repulsion, thus making it rise. Either way, says Daedalus, a convective instability should be initiated. A cloud charged by his Wimshurst balloon should rapidly develop into a rainstorm. More dramatic still, the generous initial charge on the cloud should be amplified up by the usual atmospheric electrical processes, giving a thunderstorm.

A cloud seeded by DREADCO's 'electric rainer' will release its deluge some while after being triggered, and some distance down-wind. Experience will be needed to judge the best site and timing to soak a given target; but customers should then rush to use the service. Farmers will discharge promising clouds onto their parched crops while police discourage troublesome marches and demonstrations. Sponsors of expensive sports fixtures will pre-emptively drain threatening cloud formations before they reach the sportsground. And unscrupulous evangelists may call thunder and lightning dramatically down from the heavens in support of their claims to higher authority.

David Jones

Dangers of lubricants used with condoms

SIR—Most condoms are used for contraceptive purposes. Recently, however, their use has been widely promoted to help

— showed a similar pattern, with reductions of up to 95% from initial values.

Clearly, users of latex condoms should

Mean percentage loss in tensile strength ($n=10$)

	'LUBRICANT'					
	Baby oil		Petroleum jelly		Corn oil	
	15 min	60 min	15 min	60 min	15 min	60 min
Durex (FRG)	92	91	70	69	74	68
Durex Elite (UK)	92	92	52	64	71	72
Lifestyles Nuda (UK)	73	77	51	61	64	73
Lifestyles Nuda (USA)	69	76	62	71	64	67
Mates (UK)	68	84	45	65	63	68
Ramses Extra (USA)	93	93	75	80	77	78
Trojan Extra (USA)	91	93	82	83	71	79
Duo (Netherlands)	91	90	55	63	55	80
Lifestyles Extra (UK)	76	84	35	53	39	65
Lifestyles Extra (USA)	76	84	41	67	65	70
Mates Tough (UK)	76	84	43	67	58	67

limit the spread of sexually transmitted diseases, especially AIDS¹. Many studies show that condoms are an effective barrier to the causative microorganisms².

It is widely held that petroleum- or oil-based lubricants such as baby oil, petroleum jelly or corn oil should not be used with latex condoms. Some manufacturers warn against such use on their packaging. Unable to find any published data to support this view, we tested major brands of latex condoms and samples of lubricants.

Each condom unpacked and exposed to oils at 37°C. After exposure, the tensile strength of the condoms suffered major, and often drastic, losses after a short period (see table).

Other physical properties — elongation at break, burst pressure and burst volume

not use additional lubricants that are petroleum- or vegetable-oil based, or else their condoms are very likely to fail. On the other hand, three water-based lubricant products we have tested, Duragel, Duracreme and Senselle, do not adversely affect the physical properties of condoms.

NICK WHITE

KAREN TAYLOR

ADRIAN LYSZKOWSKI

JOHN TULLETT

COLIN MORRIS

London International Group,
Research and Development, Unit 205,
Cambridge Science Park,
Cambridge CB4 4GZ, UK

1. Koop, C.E. *J. Am. med. Ass.* **256**, 2784–2789 (1986).

2. *J. Am. med. Ass.* **259**, 1925–1927 (1988).

Evolution of the early angiosperm groups

SIR—Chris Humphries (*Nature* **333**, 300–301; 1988) gives an interesting commentary on the phylogenetic classification of the angiosperm subclass Hamamelidae. He recognizes the role that the palaeontological record can play but states that the study of systematic patterns will probably be more significant for understanding the early evolution of this and other angiosperm groups.

In arguing this point, Humphries paraphrases comments that I made (*Nature* **331**, 304–305; 1988) about the nature of the fossil record of early angiosperms. In doing so, he unfortunately misrepresents me, presenting a view diametrically opposed to what I actually said. Rather than claim that the “processes by which angiosperms came to dominate low-land sediment-accumulating habitats are the most important for understanding angiosperm radiations,” I said that floras from these habitats are the only ones well represented in the fossil record and for this reason palaeontological evidence alone cannot give a comprehensive picture of the radiation of the angiosperms. This was one of my main criticisms of Lidgard and Crane's analysis of the fossil record (*Nature* **331**, 344–346; 1988) which in my view was making wide-ranging claims about early angiosperm radiation that are not justified by the nature of the data.

CHRISTOPHER J. CLEAL

37 Lipscombe Close,
Newbury, Berkshire RG14 5JW, UK

Image-processing packages

SIR—Kendall Preston (*Nature*, **333**, 611–612; 1988) makes the valid point that the number of different image-processing software packages, and the lack of standards in data format, have produced much extra work and programming for researchers. His final point concerns the problem of machine specificity, where image-processing packages are wedded to particular hardware. Here, I think, the article is misleading on current trends in image processing. The UNIX operating system and the C language are becoming very prevalent in the image-processing community, making it possible to create truly portable image-processing packages. Several are currently available, including the one I wrote and distribute, called HIPS (Landy, M.S., Cohen, Y. & Sperling, G. *Comp. Vision Graphics Image Process.* **25**, 331–347; 1984), which is used on many different machines.

MICHAEL LANDY

Department of Psychology,
New York University,
6 Washington Place,
New York, New York 10003, USA

Latex gloves not enough to exclude viruses

SIR—We have examined latex gloves from four manufacturers using scanning electron microscopy and energy dispersive X-ray analysis. All of the gloves had pits 3–15 µm wide and up to 30 µm deep on both interior and exterior surfaces. Irregular particles (30–50 µm) containing silicon and magnesium were embedded in the latex deeply enough to cause pits themselves. Freeze-fractured sections of all gloves showed cavities throughout the matrix and tortuous channels (5 µm) penetrated the entire thickness of the glove (see figure). Despite a recent report that gloves exclude virus cultures (Dalglish, A.G. & Malkousky, M. *Brit.*

J. surg. **76**, 171–172; 1988), our findings suggest that double gloving, possibly supplemented by a surfactant/viricide, is a prudent expedient for those handling HIV or hepatitis B virus-infected material.

SUSAN G. ARNOLD

JAMES E. WHITMAN JR

Advanced Biotechnologies,
2150 Tech Road, Silver Spring,
Madison 20904, USA

CECIL H. FOX

MICHELE H. COTTLER-FOX

National Institutes of Health,
Bethesda, Maryland 20892
and
Department of Medicine,
Georgetown University Hospital,
Washington, DC 20057, USA



Cross section through a latex glove ($\times 1,700$) with the exterior surface on the left.

Identification of virus causing recent seal deaths

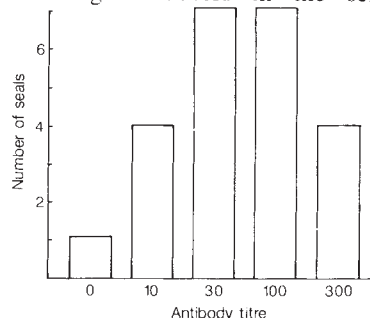
SIR—We recently reported¹ the isolation of a herpesvirus and a virus tentatively classified as a picornavirus, from organs of harbour seals (*Phoca vitulina*), which had died during recent outbreaks of acute disease with high mortality in the North and Baltic seas. We suggested that further evidence for the causative role of these viruses for the outbreaks should come from serological studies on samples collected from animals during and after the outbreaks.

In virus-neutralization assays, we now show that, although virus-neutralizing antibodies against both viruses are present in sera from harbour seals in endemic areas, there is no correlation between the occurrence of disease symptoms and the development or rise of antibody titres against either virus. We also show that the immunization of young harbour seals in the seal orphanage in Pieterburen, The Netherlands, with inactivated preparations of either of these viruses, which did elicit virus neutralizing antibodies, fail to protect the animals against fatal disease (manuscript in preparation). These data suggest that these two virus infections, rather than being the primary cause of the disease outbreaks, are opportunistic infections occurring in animals suffering from another disease.

In a further attempt to identify the primary cause of the outbreaks, which on the basis of epizootiological observations is generally believed to be infectious², we extended our serological studies to other viruses of carnivores known to cause similar disease symptoms. For this reason, and as it had been noted by Drs A. Bergman and B. Klingeborn (National Veterinary Institute, Uppsala, Sweden) that the postmortem findings observed are often similar to those of canine distemper virus (CDV) infections in dogs, we carried out CDV serology on serum samples collected from seals in The Netherlands, Denmark, Germany, Sweden and the United Kingdom. We tested all the samples in the virus-neutralization test described³ previously (see table). We tested 21 serum samples from harbour seals collected in 1984 in the seal orphanage in Pieterburen after the outbreak of acute disease caused

by Phocid herpesvirus 1⁴. All these samples are negative in this assay (titres <10). Also, 16 samples collected from seals in March 1988 in the orphanage just before the start of the outbreaks were negative.

From May 1988, the first orphan seals born this year were brought into the orphanage and kept in separate groups of 10 to 15 animals. We force-fed these animals to exclude lactogenic transmission of antibodies. Once the first symptoms started to occur, we detected CDV-neutralizing antibodies in the serum



Development of CDV-neutralizing antibody titres within 14 days in orphan seals seronegative at admission to the Pieterburen orphanage.

samples. Of 23 baby seals brought to the orphanage between 3 July and 8 August clinically healthy and without CDV-neutralizing serum antibodies, 22 developed CDV neutralizing antibodies with titres of 10–300 within 14 days, and they all developed disease symptoms in this period (see figure).

Most of the baby seals brought to the orphanage after 9 August showed disease symptoms and were seropositive on arrival. Serum samples collected from grey seals (*Halichoerus gryphus*), which had been in the orphanage during the outbreak and survived the disease, showed CDV-neutralizing antibody titres ranging from 10,000 to 30,000. From one of these animals, a serum sample collected before the outbreak in March 1988 was available and this proved to be negative. Of seven serum samples collected from harbour seals in Denmark which had died with acute symptoms in May 1988, two showed antibody titres of 10 and 1,000 respectively. Of 35 samples collected in July and August 1988 from harbour seals in

Germany at different stages of the disease, 23 were seropositive (titres 10–30,000; $\bar{x}$ =100). Eight serum samples from harbour seals suffering from the disease in Sweden in August 1988 were tested. All except the serum from one baby seal show CDV-neutralizing antibody titres ranging from 10 to 30,000 ($\bar{x}$ =138). Finally, all eight serum samples from harbour seals in a seal sanctuary in the United Kingdom which had survived the disease show CDV-neutralizing antibody titres ranging from 30 to 3,000 ($\bar{x}$ =363).

These serological data clearly show that an infection of CDV, or a closely related morbillivirus, occurred in the respective seal populations after April 1988. That the antibodies found were not directed against measles virus (MV), a closely related morbillivirus⁵ was shown in an MV-specific haemagglutination-inhibition assay⁶ in which CDV neutralizing seal sera were negative. That a few samples were not positive may be because they were taken mainly during the acute stage of the disease when no anti-CDV antibodies had yet developed. The results obtained with the paired serum samples from individual animals in the orphanage show that the infection coincided with the occurrence of the disease symptoms. Furthermore, the clinical symptoms observed during the outbreaks, which affected respiratory, gastrointestinal, central-nervous and cutaneous systems, are similar to those of canine distemper, which is also frequently accompanied by secondary viral and bacterial infections⁵.

We therefore conclude that the primary cause of the disease outbreaks of seals is infection with CDV or a closely related morbillivirus. The failure so far to isolate the virus from organs of affected animals may be explained by the fact that no isolation techniques were used which favour the isolation of CDV⁵. We have now started *in vivo* and *in vitro* virus-isolation procedures to compare properties of the virus with those of CDV isolates from dogs. Because in seal sanctuaries there is an urgent need for a preventive vaccine, and the use of live vaccines is in general inadvisable for wild animals, we have now started to evaluate in seals the value of a subunit iscom CDV vaccine, recently shown to be effective in dogs⁷.

A. D. M. E. OSTERHAUS

National Institute of Public Health and
Environmental Protection,
Bilthoven, The Netherlands

E. J. VEDDER

Seal Orphanage, Pieterburen,
The Netherlands

		CDV-neutralizing serum antibody titres ≥ 10 in harbour seals.						
		Year	1984		1988			
		Month	M	A	M	J	J	A
Clinical symptoms								
The Netherlands* (Pieterburen orphanage)	(–)†		0/21‡	0/16	0/5			
	(+)					1/14	8/33	8/17
						0/1	20/21	43/52
Denmark	(+)				2/7			
Germany	(–)						2/10	3/4
	(+)					10/12	10/10	
Sweden	(+)							7/8
United Kingdom	(+)							8/8

*Serum samples obtained from Drs E. Vedder (NL), P. Grauballe (D), E. Vedder (D), B. Klingeborn (S) and S. Anderson (UK). †(–): Clinically healthy, (+): Clinically ill. ‡Number positive/number tested.

- Osterhaus, A. D. M. E. *Nature* **334**, 301–302 (1988).
- Statement of the emergency working conference on seal deaths in the Baltic and North Seas, 11 August 1988, London.
- De Vries, P., Uytendaele, F. G. C. M. & Osterhaus, A. D. M. E. *J. gen. Virol.* **69**, 2071–2083 (1988).
- Osterhaus, A. D. M. E. *et al.* *Archs Virol.* **86**, 239–251 (1985).
- Appel, M. J. G. in *Virus Infections of Vertebrates* (ed. Horzinek, M. C.) 1 (Elsevier, Amsterdam, 1987).

The cost of complexity

R. W. Chapman

The Collapse of Complex Societies. By Joseph A. Tainter. Cambridge University Press: 1988. Pp.250. £27.50, \$39.50.

The Collapse of Ancient States and Civilizations. Edited by Norman Yoffee and George L. Cowgill. University of Arizona Press: 1988. Pp.333. \$35.

DURING the nineteenth century, the discovery and study of 'lost civilizations' gave great impetus to the development of archaeology. The surviving traces of such civilizations — whether in the Mycenaean and Minoan palaces of the Aegean, the frontier defences of the Roman Empire, the pyramids of the Nile Valley, the ziggurats of Mesopotamia or the ceremonial centres of the Maya in the dense jungles of Mesoamerica — provided powerful images for the recently industrialized states of Western Europe. The predominant image, and perhaps also a warning, was of the decline, decay and collapse of cultures. The archaeological evidence testified to the existence of complex societies, with highly centralized government, comparatively dense populations (sometimes in urban centres), stratified social organization, specialist workers and the exercise of armed force by the ruling classes. If these state societies with their expanding empires had been unable to sustain themselves, and had succumbed to rapid collapse, what had caused this decline? And was there a lesson to be learnt by the modern Western world?

Archaeologists and other social scientists have continued to debate the answers to these questions during the present century, but it is fair to say that recent interest has been more in the reasons for the *emergence*, rather than the *collapse*, of early states. Much of the attention given to the topic has arisen directly out of advances in our knowledge of the demography and subsistence and settlement patterns of early states, which have stemmed from modern archaeological surveys and excavations. For example, the impact of fieldwork on our understanding of the origins of the state in Mesoamerica can be seen in Blanton *et al.*'s *Ancient Mesoamerica* (Cambridge University Press, 1981), while McC. Adams (*The Heartland of Cities*; University of Chicago Press, 1981) has demonstrated the influence that arch-

aeology has had on Mesopotamian studies. Cross-cultural comparison of the evolution of these, and other early civilizations, is now becoming more reliable — see 'The Evolution of Civilizations' by Henry Wright in *American Archaeology: Past and Future*, edited by Meltzer *et al.* (Smithsonian Institution Press, 1986).

But what of the collapse of these civilizations? What new data can be brought to

and the debate between conflict and integration theories over their emergence. Then he proceeds to classify theories for state collapse into eleven types, such as resource depletion, catastrophes, insufficient response to changing circumstances, intruders, internal conflict or mismanagement, economic explanations, and what he calls 'mystical factors' (the last of these are exemplified in the works of Spengler and Toynbee). He finds all these theories wanting, because they cannot account for empirical variability, or are empirically untestable, or because they underestimate the capacity of early states to respond to stresses upon them.

Tainter's preference is for an economic model, through which the costs and benefits of complex societies can be examined, and which is based on the argument that all states reach a point at which costs outweigh benefits, so that for every new 'investment' in complexity, there is a diminishing 'marginal return'. He supports the model by reference to agricultural production, information processing, sociopolitical control and specialization, and overall economic productivity, using data as diverse as traditional subsistence agriculture, contemporary spending on higher education in the United States, and the growth of bureaucracy in the modern British navy! According to this model, there comes a point at which, unless there is a fresh investment in energy sources or technical innovation, disintegration of the state becomes more attractive to the members of that state than further investment in complexity. Tainter assesses the model's usefulness in three case-studies of collapse, for the Western Roman Empire, the classic Maya state and the Chacoan society of north-west New Mexico. His conclusion is that it does help us to understand these examples, and that such collapses were not disasters or catastrophes, but 'economizing processes', and may not have been

at all disastrous to most of the population.

In contrast to Tainter, Yoffee and Cowgill have opted for a more detailed examination of a small number of case-studies of collapse (two papers on Mesoamerica, two on Mesopotamia, one on the Roman Empire and one on China) and a more pluralistic approach to theory. The case-studies are viewed in the light of both archaeological and documentary sources, the coverage is somewhat uneven (for example, the paper on the Roman Empire



All fall down — the ruin of Baalbek, an illustration from *The Holy Land* by David Roberts (1843).

bear on the problem? How successfully can we evaluate competing explanations for collapse? Does the problem have the same, forceful relevance for us as it did for our nineteenth-century predecessors?

Tainter, and the contributors to Yoffee and Cowgill's book, attempt to answer these questions, but in different ways. Tainter's concern is with the search for a general explanation, stemming from a body of economic theory. First he summarizes the characteristics of early states

From *The Collapse of Complex Societies*

is very short and makes little reference to archaeology) and in some cases the authors seem more accustomed to writing about states than about their collapse. The attention to detail exposes some weaknesses of theory: thus Millon's discussion of the end of the great city of Teotihuacan, in the valley of Mexico, is very speculative, while Bronson's treatment of the role of 'barbarians' in the fall of states would have benefited from an analysis of proximate as opposed to ultimate causality.

Like Tainter, the authors in the edited volume concentrate on collapse as a process of political fragmentation. There is a comparable dissatisfaction with the empirically untestable 'models' of Spengler and Toynbee. Further similarities can be seen in the recognition that state societies are atypical in human evolution, being a very recent phenomenon, and that we have overestimated the degree to which they were self-regulated. Although none of the contributors proposes the same model for collapse as Tainter, it is clear that the relationship between the costs and benefits of complexity is central to the thinking of at least two of them. Thus Yoffee writes that "stability in historic states and civilizations is maintained when the periphery considers that the resources it provides the center also return benefits to itself", while Cowgill considers that one of the main problems faced by large states is how they managed to "balance their costs and expenses when they could not simply capture resources from others".

Differences of emphasis are also clear between the two books. There is a stronger condemnation of biological analogies for cultural evolution in Yoffee and Cowgill (see Yoffee's introduction), and McC. Adams argues that there is no equivalent of extinction in human cultures. Indeed the same author goes further to argue strongly that collapse of states and civilizations is neither absolute, nor inevitable and determinate. We should not treat states as 'stable', but rather consider, as a problem worthy of study, how they cope with the problems of short-term instability. Once again, the equilibrium model of complex cultures is seen as inappropriate.

Perhaps the clearest comparison between the two books can be made by reading Cowgill's excellent discussion of the ways in which, in the social sciences, we conceive of problems such as collapse, and how we attempt to explain these problems. Cowgill contrasts general, highly abstract theories with particular, detailed analyses of individual case-studies. He prefers 'scientific' explanation, but argues that we need to develop "models that achieve high outputs of usefully accurate predictions and postdictions of significant phenomena in return for relatively economical inputs of relevant data and parsimonious theory". What would be fascinat-

ing would be to hear his views on Tainter's 'explanation' of collapse, which, I suspect, he would characterize as being too abstract to account for variation in particular observations. How do states vary in rates of collapse, how far do other institutions fall apart along with political fragmentation, how far does collapse have a deleterious effect on central and more peripheral populations, why was collapse beneficial to some peripheral populations and not to others, and so on? These are the kinds of questions with which Tainter is less concerned, but any theoretical approach which cannot accommodate them will be seen as less than successful.

Are there any lessons to be learnt from these two books? Clearly the ways in which we conceive of early states and their collapse have changed, at least in part because of the input of new data from archaeology. Nineteenth-century views of

collapse were over-dramatic — they overestimated the stability of early states, they employed what we would now see as rather simplistic, biological analogies for the rise and fall of civilizations, and the data on which they were based were restricted essentially to the main political centres. Archaeology, in alliance with other social sciences, is now working with an infinitely larger database and using different theoretical approaches, enabling, for example, a comparative analysis of change in political and other cultural institutions. Our very conception of collapse, let alone its explanation, is sufficiently different as to nail, once and for all, gloom and doom analogies drawn from the past to predict the future. □

R.W. Chapman is a Lecturer in the Department of Archaeology, University of Reading, Whiteknights, Reading RG6 2AA, UK.

Clarifying chaos

Michael Berry

Mathematics and the Unexpected. By Ivar Ekeland. *University of Chicago Press: 1988. Pp.146. \$19.95, £15.95.*

THIS century has seen three revolutions in the science of dynamics. The first two were relativity and quantum mechanics, which changed the laws of physics under conditions of speed and size far removed from our direct experience. The third is chaology, which left the laws of motion (Newton's) unaffected but has radically altered our understanding of the behaviour they describe. It is this revolution that is one of the two themes of this book.

Ekeland's treatment is both philosophical and historical. In the first part of the book he describes the old newtonian-laplacian view of mechanics, in which the reversibility and uniqueness of solutions of the equations of motion were taken to imply that events (abstracted as orbits of bodies under forces) are predictable, and "Past and future are seen as equivalent, since both can be read from the present. Mathematics travels back in time as easily as a wanderer walks up a frozen river".

The shattering of this interpretation by Poincaré's discoveries and their recent extensions and widespread applications is discussed in the second and central part of the book. By carefully chosen examples, in which the mathematics is explained lucidly and with the minimum of technicality, the main message is reached: "... a purely deterministic law may materialize in a totally random series of observations if part of the information is withheld, as it must be in any practical situation. ... Like the queen of England, determinism reigns but does not govern".

There are some errors. It is not the case that "Dissipative systems cannot have complicated trajectories": when forced, they can, and the trajectories explore the 'strange attractors' whose most familiar manifestation is the unpredictability of the weather. It is not true that the appearance of the Feigenbaum constant "in many different circumstances has been one of the great scientific puzzles of past years": Feigenbaum's discovery of the constant was accompanied by his mathematical explanation of its universality. And Arnold's cat is confused with the baker's transformation, which is a different chaotic mapping. These mistakes mar but do not ruin the exposition. They are more than compensated for by an elegance and directness of expression.

My enthusiasm does not extend to Ekeland's treatment of catastrophe theory, which is his second theme. Thom's grand vision of a library of nature's forms is explained clearly enough apart from minor technical errors, but the discussion is a decade out of date. It is no longer true that "there has not been a single undisputed success of catastrophe theory in the field of experimental science". A flourishing new branch of optics has been created, which has furnished quantitative explanations of many natural phenomena and stimulated a variety of experiments. The author seems unaware of this, and of Arnold's enormous extension of Thom's classification (far beyond the seven dimensions at which Ekeland asserts that the theory fails).

This is a Jekyll and Hyde of a book, on which the verdict has to be that the chaology deserves applause but the catastrophe theory is inadequate. □

Michael Berry is a Professor in the Department of Physics, University of Bristol, H.H. Wills Physics Laboratory, Tyndall Avenue, Bristol BS8 1TL, UK.

Events frozen in time

T.M.L. Wigley

The Little Ice Age. By Jean M. Grove. Methuen: 1988. Pp. 498p. £85, \$144.

IN THIS outstanding piece of scholarship, Jean Grove has succeeded in providing an immensely detailed, critical account of climatic changes without losing sight of the global perspective in which these changes are most usefully viewed. Don't be misled by the title. There is much more to this book than just an account of the Little Ice Age; indeed, I know of no better review of the climate of the Holocene (the past 10,000 years).

The Little Ice Age is the most important climatic fluctuation that has occurred in historical times. Although there are variations in the defined time limits, it can broadly be assumed to have begun in Mediaeval times and ended in the nineteenth century. During this period, Europe (from which the most voluminous evidence comes) was generally colder than today, Alpine glaciers were far advanced, and various tales of the woes inflicted by particularly severe years or groups of years may be found in historical sources. These details are brought out vividly in the text, in words, immaculate diagrams and pictures. But was this a world-wide event? Grove states "It is always tempting to seek universals, and they are easier to promote when the data are imprecise . . .". By carefully reviewing the evidence of glacier fluctuations throughout the world, never losing sight of the attendant uncertainties, she concludes that the event was global in scale, albeit with regionally specific variations superimposed. Grove's conclusion corresponds to that reached by other authors, but the information has not previously been considered so comprehensively.

The next leading question concerns the uniqueness of the Little Ice Age. For many years it has been known that the Holocene was punctuated by similar events in specific regions, but the perceived wisdom has been that few of these were of global scale. The general picture of the Holocene is of an early warm period followed by overall cooling with more pronounced cool episodes every 2,000–3,000 years (the Little Ice Age being the last of these). This picture has unfortunately been perpetuated in a number of authoritative reviews of climatic change. Grove is more realistic and more sceptical, correctly pointing out, for example, that there is "no certainty that there ever was . . . a single identifiable warm period in the [early] Holocene". Instead, what emerges from her analysis is a Holocene climatic history characterized by a sequence of ten or more global-scale Little Ice Ages, fairly irregularly spaced and each lasting a few centuries. As Grove notes, this was the view espoused by Matthes when he introduced the term 'Little Ice Age' in 1939.

The causes of these periods of glacial advance are difficult to ascertain. The most popular candidates are changes in volcanic activity or in solar irradiance, most probably modified by induced changes in albedo and/or ocean circulation. Grove leans slightly towards the solar explanation. I agree, and have been able to support this possibility statistically using carbon-14 data which appeared since Grove's book was written.

Why is the Little Ice Age (and the earlier similar events) so important? The future climate of the globe may well be dominated by the influences of man-made greenhouse gases such as carbon dioxide. But whatever these effects are (and there are considerable uncertainties in this regard) they must be judged against the natural climatic variability that occurs on the same time scale, decades to centuries. The Little Ice Ages documented and described in this book are this natural varia-

Reviews on IMMUNOASSAY TECHNOLOGY

Edited by S. B. Pal

Immunoassay techniques are continually being developed and refined, and applied in a broader range of uses. The books in this series comprise review articles and descriptions of methods of immunoassay, which are not reliant on radioisotopes, and reflect the growing importance of immunoassay technologies.

Volume 1.

October 1988 £37.50 256pp
234x156mm 0 333 452852

Volume 2.

October 1988 £37.50 216pp
234x156mm 0 333 46849X

Order these books through your bookseller or send your order to:

**Richard Gedye, Macmillan Press Ltd,
Houndmills, Basingstoke RG21 2XS,
UK.**

bility. Grove estimates that these events correspond to global coolings of 1–2° C (in summer at least), although I would prefer slightly lower numbers. But even 1–2° C is appreciably smaller than most projections of global warming between now and the mid-twenty-first century.

I can find little of consequence to fault in this book. The author herself, however, has one regret and this is that she completed her work on it too soon to be able to make full use of the equally impressive work of Friedrich Röthlisberger (*10,000 Jahre Gletschergeschichte der Erde*; Verlag Sauerländer, Aarau, 1986). In my view, her own study is just as valuable and is certainly broader in scope. She makes a great deal of the foreign literature, including much of Röthlisberger's work, accessible to the English-speaking scientific community. In the end she reaches the same conclusions as Röthlisberger, but perhaps in a more balanced way.

As future climatic change becomes an increasingly important issue, we may tend to lose sight of the past. The literature, even on the Holocene, is extensive and diverse, spanning many disciplines. For those eager to gain an appreciation of the most relevant part of our climatic heritage, this book should be compulsory reading.

T.M.L. Wigley is a Professor in the Climatic Research Unit, University of East Anglia, Norwich NR4 7TJ, UK.



A.T. Grove

Slippery slope — Holocene and Little Ice Age moraines in front of a glacier in western China.

Soviet-style cell motility

Dennis Bray

Cytoskeleton. By Alexander D. Bershadsky and Juri M. Vasiliev. *Plenum: 1988.* Pp.298. \$47.40, £27.90.

SCIENCE monographs have become such a stylized feature of Anglo-American culture that it seems strange to find one of the genre emerging from the Soviet Union. Indeed, this glossy volume, written in impeccable English and full of up-to-date references and micrographs, is almost uncannily good — as though the two Muscovites had taken centre stage in a Country and Western show with a twangy rendition of “Mommas, don’t let your babies grow up to be cowboys”. Where did they learn to give such a polished performance? How did they — with all the difficulties that are said to exist in the Soviet Union in obtaining Western periodicals and travelling to conferences — become so well informed? Whatever the answer, this volume sends a clear signal that at least one active and critical school of cell biology has survived the Siberian winter. May it flourish now that Spring is on its way!

This is the best book on the cytoskeleton yet written. To be sure, one expects improvements as the field matures; as the flurry of newly discovered cytoskeletal proteins begins to settle, we expect to discern the contours of a larger structure beneath. But the authors have not missed any opportunities in the integration of their material. A good half of the book is devoted to the organization of cytoskeletal elements — links between different sets of filaments are examined as well as their association with the plasma membrane and the cell nucleus. One chapter deals with the interplay of cytoskeletal elements during cell locomotion, another with cell division. The final chapter on the neoplastic transformation of fibroblasts, a research interest of the authors, concerns the fascinating links between oncogenes and the cytoskeleton.

Most of the important new developments have made it into the book. Dynamic instability and kinesin, integrin and talin, intermediate filament sequences, control over tubulin mRNA stability — all are mentioned, although some very briefly. The text is systematically organized, with each filament system being introduced in turn and each cytoskeletal protein or mechanism or drug receiving its allocation of space. Areas of uncertainty or argument (of which, in this field, there are a large number) are dissected with scrupulous impartiality: “Here are the possible mechanisms, A, B, and

C”, the authors seem to say, “And here is the evidence for each of them, in turn”. It is a scientific style for a scientific book; appropriate, I suppose, but the result is a little dry and joyless. This is probably not a volume to capture a young person’s imagination.

But then the book was not written as a student text on cell movements. It is, explicitly, a monograph in a series on cellular organelles, the organelle in question being the cytoskeleton. As such, it per-

forms its function to an exemplary degree, with many useful tables and a wealth of original, somewhat idiosyncratic, line drawings. Complete, accurate and insightful, it will be a valuable resource for research workers and a useful supplementary text for courses in cell biology. I hope it does well. □

Dennis Bray is in the Medical Research Council Cell Biophysics Unit, King’s College London (KQC), 26–29 Drury Lane, London WC2B 5RL, UK.

Portrait of a cell

John Sommerville

Paramecium. Edited by H.-D. Görtz. *Springer-Verlag: 1988.* Pp.444. DM248, £83.

TO DEVOTE a whole volume to a single eukaryotic cell may seem an extravagance. But not in this instance, for *Paramecium* is one of our favourite organisms.

Superficially familiar to students as a busy little forager spiralling through pondwater, *Paramecium* has much to offer the modern research worker. Indeed, the ciliated protists as a group are among the most fascinating of all cells, having evolved to a level of complexity rarely seen elsewhere. Complexity is evident not only in their intricate structural features but also in their genetic mechanisms, metabolic processes and locomotor activities. Thus many of the features of higher organisms — differentiation, responsive behaviour, a digestive system — are seen here in a single cell, and the endeavours to understand them, as manifested in *Paramecium*, are reviewed in this volume.

Although *Paramecium* is accessible to a wide variety of experimental approaches, progress has been limited by some technical problems. One is the difficulty associated with cell culture in axenic media (*Paramecium* eats bacteria in nature); another is the dearth of suitable (micronuclear) mutations for genetic analysis. Happily, progress in both of these areas is reported in the book.

These and other problems and prospects with using *Paramecium* as an experimental organism are identified by John Preer, whose foreword provides a useful overview of various studies described in the text. Also, an early chapter by Richard Allen gives the background to the ultrastructural aspects of the systems which are tackled in more detail later on. Each system (ranging through taxonomy, mating types, cell cycle, nuclei, immobilization antigens, mitochondria, ion channels, cilia, chemokinesis, lysosomes, trichocysts, cytoskeleton, endosymbionts, ecology) is the subject of a separate contribution written by appropriate specialists. Although the 24 chapters vary in length and detail of content, each is clearly writ-

ten and amply referenced. Related topics are grouped together, and format is consistent throughout, resulting overall in a coherent, well-integrated volume.

Yet several aspects of *Paramecium* remain enigmatic or involve biological principles which we are only beginning to understand. These aspects transcend the boundaries of individual chapters. For instance, nuclear dimorphism, whereby intact copies of the genome are preserved in micronuclei, and amplified, fragmented, deleted and rearranged copies are expressed in macronuclei, provides a model not only for germ line versus soma effects (Chapter 8) but also for chromatin activation (Chapter 10) and the nature of ageing (Chapter 9). Differentiation and inheritance of cortical structures (Chapters 16, 20 and 21) relate to the construction and reproduction of complex supramolecular assemblies by mechanisms not yet understood. The relationship between excitable membranes (Chapter 13), ion-channels (Chapter 15), motility (Chapter 14) and locomotor behaviour (Chapter 18) become clearer through the analysis of behavioural mutants (Chapters 15 and 17). Patterns of cytoplasmic inheritance, through mitochondria (Chapter 12) and bacterial endosymbionts (Chapters 22–24), or through less well defined ‘cytoplasmic states’ which can control macronuclear gene expression (foreword and Chapters 4 and 11), involve interactions between different genomes.

Paramecium presents us with a rich canvas. Although this is not the only book available on the organism, or on ciliates in general, it does contain a comprehensive coverage of recent advances and illustrates that the details on that canvas are only just beginning to be filled in. □

John Sommerville is in the Department of Biology and Preclinical Medicine, University of St Andrews, St Andrews, Fife KY16 9TS, UK.

New in paperback

- *Forever Undecided: A Puzzle Guide to Gödel* by R. Smullyan. Publisher is Oxford University Press, price is £5.95. For review see *Nature* 328, 212 (1987).
- *Speakable and Unspeakable in Quantum Mechanics* by J.S. Bell. Publisher is Cambridge University Press, price is £8.95, \$14.95. For review see *Nature* 331, 667 (1988).
- *Was Einstein Right?* by C.M. Will. Publisher is Oxford University Press, price is £5.95. For review see *Nature* 324, 191 (1986).

The molecular genetics of embryonic pattern formation in *Drosophila*

P. W. Ingham

Molecular Embryology Laboratory, ICRF Developmental Biology Unit, Department of Zoology, South Parks Road, Oxford, OX1 3PS, UK.

Analysis of the genes that control the early events of Drosophila embryogenesis is providing details of the molecular processes underlying the positional specification of cells. There are two distinct phases: the first precedes the cellularization of the blastoderm embryo and is associated with a cascade of interactions between transcriptional regulators; the second occurs after cellularization and depends on communication between cells. These processes may be conserved in a wide range of invertebrates and vertebrates.

THE quest for substances that determine the fates of embryonic cells first began at the end of the last century¹. Now we have a plethora of such molecules, principally those controlling the development of the dipteran *Drosophila melanogaster*. Behind this success lies the intellectual inspiration of the geneticist E. B. Lewis. It was his innovative analysis of the *bithorax* homeotic mutations^{2,3} that established the paradigm on which the contemporary analysis of *Drosophila* development has been based. The veracity of Lewis' notion of families of structurally related genes controlling the specification of the body plan has underpinned the strategies adopted in the molecular cloning and characterization of many of the developmental genes in *Drosophila*⁴⁻⁶. The successes of this approach now leave us on the threshold of a new era in the study of development. Here I review our current knowledge of the ways in which these genes interact to control the early events of embryogenesis.

Insect embryos have been a popular subject of analysis among embryologists for some time. These studies⁷ have generated the fundamental concepts on which our current understanding of the genetic control network is based. Insect bodies are metameric: that is, they are composed of serially repeating units, or body segments, which differentiate into particular structures and patterns according to their position (see Fig. 1). The generation and diversification of segments during embryogenesis depends on two distinct but integrated processes: the subdivision of the embryo into reiterated units, and the specification of their pathways of differentiation. In many insects segments are generated sequentially as the cells of the embryo proliferate. By contrast, in *Drosophila* and other so-called long germ-band insects, the entire body plan is established simultaneously at the blastoderm stage of embryogenesis.

The blastoderm is a single-layered epithelium formed by the migration to the periphery of the embryo of the majority of the nuclei generated by the initial syncytial cleavage divisions^{8,9}. The embryo remains a syncytium while the nuclei undergo three further divisions following this migration. Membranes then grow down from the cortex to separate each nucleus, so forming the cellular blastoderm. By the completion of this process, the fate of each newly formed cell is already well established¹⁰, although the extent to which its developmental potential is determined varies with its position. For instance, cells are restricted to contributing to only one body segment^{11,12}, though the part of a segment to which they contribute will depend on later interactions between their progeny. Similarly, cells are already restricted to contributing to either mesoderm or ectoderm, but the decision to follow the neuroectodermal pathway will depend on later events¹⁰.

The cellular blastoderm can thus be viewed as providing a framework for the subsequent elaboration of the final body

pattern¹³. Experiments involving the temporary separation of the egg and/or the transplantation of cytoplasm from one part of the egg to another demonstrated that the antero-posterior body plan is generated progressively during the early stages of embryogenesis, under the influence of organizing centres localized at each pole of the embryo. These manipulations, originally performed by Sander and colleagues⁷ on leafhopper and midge embryos, have more recently been repeated in *Drosophila*^{14,15}. In addition, an extensive mutational analysis has been undertaken¹⁵ which has resulted in the identification of several genes involved in the synthesis and localization of these sources. Absence of any of these genes from the maternal parent of an embryo has dramatic consequences for the final pattern, causing the gross respecification of a large portion of the body. I first consider the nature of these genes and their products, and will describe how the information they encode is translated into the fine-grain pattern of the *Drosophila* larva.

Pattern along the antero-posterior axis

A major determinant of antero-posterior pattern is the product of the *bicoid* (*bcd*) gene¹⁶. Like many other maternally expressed genes, *bcd* is transcribed in the ovary in specialized cells which form a cluster around the future anterior pole of the developing oocyte^{4,17}. The *bcd* RNA synthesized in these so-called nurse cells passes into the oocyte through cytoplasmic canals. As it enters the oocyte it is trapped, apparently by components of the cytoskeleton encoded by the genes *swallow* and *exuperantia*¹⁷. The *bcd* RNA thus becomes localized at the anterior pole of the oocyte by virtue of its site of entry. At about the same time, several other genes that are members of the *oskar* group¹⁸, are involved in the synthesis and deposition of products that also become localized, in this case at the posterior pole of the egg. Thus when the mature egg is laid, it is both morphologically and molecularly polarized. The fundamental role of these gene products in pattern specification is illustrated by their mutant phenotypes. Females lacking a functional copy of the *bcd* gene produce eggs that develop into embryos with no head or thorax¹⁶, reminiscent of those produced by removal of material from the anterior pole of a normal egg¹⁵ (Fig. 1). Embryos produced by females mutant for any of the *osk* group genes by contrast develop normal head and thoracic segments, but lack the entire abdomen¹⁷.

Localized gene products as signalling centres

The pattern-forming process continues on fertilization with the translation and translocation of information encoded by the localized RNAs. As the *bcd* RNA is translated, the protein diffuses away from the anterior pole so that it becomes gradually distributed over about one-half of the length of the egg^{19,20}

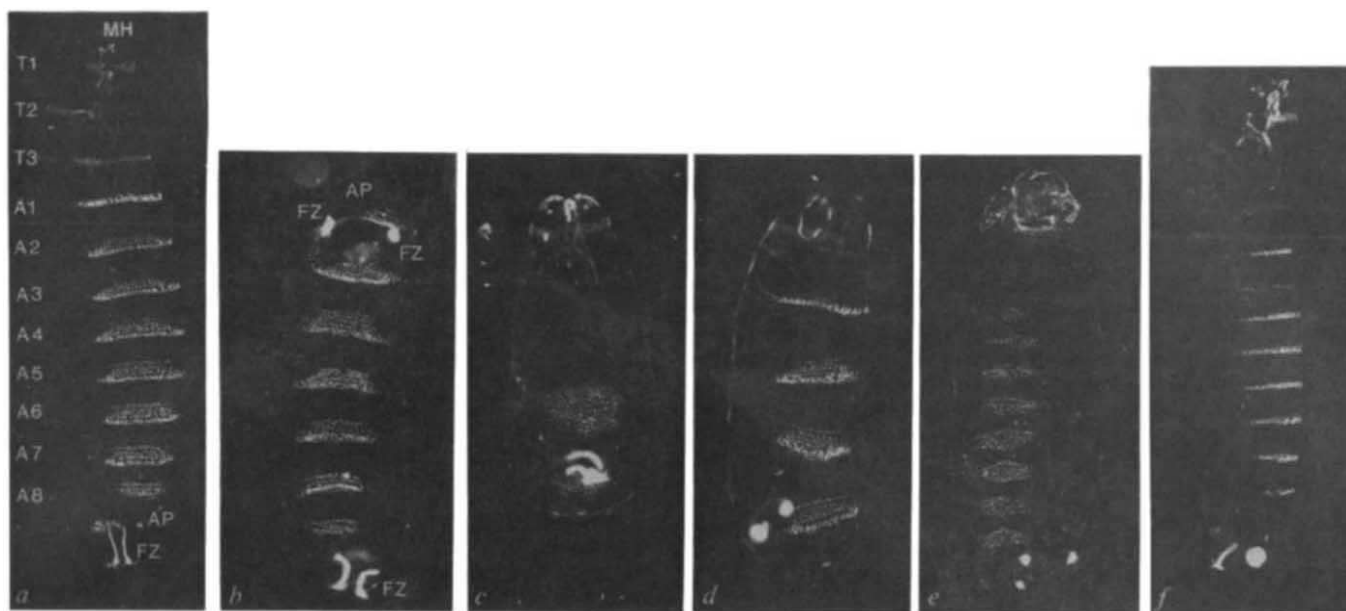


Fig. 1 *a*, The ventral cuticular pattern of a newly hatched *Drosophila* larva. Each of the thoracic and abdominal segments differentiates a characteristic belt of cuticular processes or denticles which appear white when viewed with dark-field illumination. The denticles in the thoracic segments (T1–T3) are much smaller than those of the abdominal segments (A1–A8). The morphology of each belt of denticles is unique, thereby allowing the identification of specific segments in mutant larvae (see panels *b–f*). The head segments are involution and give rise to a variety of structures, including the mouth hooks (MH). At the posterior of the larva are the anal pads (AP) and filzkörper (FZ), specialized structures of the tracheal system. *b–f*, Examples of mutations in the different classes of genes which control the pattern along the antero-posterior axis. *b*, Larva derived from a female homozygous for the *bcd* mutation. Note the absence of head and thoracic structures: these are replaced by anal pads and filzkörper typical of the posterior terminalia. *c*, Larva homozygous for the gap class mutation *kni*. In contrast to *bcd*, such mutations develop a normal head and thorax, but lack most of the abdominal segments. This phenotype is similar to that of larvae derived from females mutant for the *osk* class of genes. Note that although the abdomen is largely eliminated, the posterior terminalia (telson) develops normally. The specification of these structures depends on the maternally expressed *torso* class of genes (see legend to Fig. 2). *d*, Larva homozygous for the *h* mutation, showing the phenotype typical of the pair-rule class of mutations. In the absence of *h* activity, the denticle belts of segments T1, T3, A2, A4, A6 and A8 fail to form, the larva being composed of only half the normal number of segmental units. *e*, Larva homozygous for the segment polarity mutation *ptc*. The larva possesses the normal number of segments, but, as with all segment polarity mutations, a particular part of each segment is eliminated. In the case of *ptc*, it is the middle region of each segment (including the posterior part of each abdominal denticle belt). This is replaced with a duplication of the remaining part (the anterior denticle belt and naked cuticle) in reverse orientation. *f*, Homoeotic transformation of larval segments. Homoeotic mutations typically cause the transformation, but not the elimination, of entire segments. This larva is homozygous for a mutation in the *abd-A* gene of the BX-C. This results in all of the abdominal segments differentiating the pattern typical of the normal A1 segment (compare with panel *a*).

(Fig. 2). The protein has a putative DNA binding homoeo domain and, consistent with this structure, is found associated with the nuclei of the syncytial blastoderm. Simultaneously, information encoded and localized at the posterior pole during oogenesis by members of the *osk* group begins to move anteriorly. The details of this translocation process and the molecular form of the translocated material are at present unclear; however, it is known from transplantation experiments¹⁸ that the material is required in the presumptive abdominal region of the embryo, and not at the posterior pole itself. One member of the *osk* group, *pumilio* (*pum*), seems to be involved in the translocation of the material from the pole to its site of action²¹, whereas other members of the group are involved in its synthesis and export from the nurse cells to the oocyte.

By the beginning of the syncytial blastoderm stage there is thus an inhomogeneous distribution of at least two different maternally encoded products along the antero-posterior axis of the embryo. This quantitative information is now transformed into qualitative differences in the form of region-specific gene expression, a process requiring the interaction between the maternally derived products and the zygotic genome.

Primary subdivision of the embryo

The zygotic genes in question are members of the gap class²², so called because of their mutant phenotypes which consist of the elimination of particular regions of the body, thereby creat-

ing a gap in the antero-posterior pattern (Fig. 1). There are three principal members of this class, *hunchback*, *Krüppel* and *knirps*²², each of which is characterized by the presence of DNA-binding finger domains in their encoded proteins (refs 23 and 24; H. Jäckle, personal communication). All three genes are first zygotically expressed in the syncytial blastoderm, two division cycles before cellularization. Expression of *hunchback* is restricted to two broad domains^{24,25}, one extending from the anterior pole to 50% egg length (EL), the other from the posterior pole to about 25% EL. Similarly, *Kr* is initially expressed in a single broad band in the middle of the embryo^{26,27} (Figs 1 and 2), whereas *kni* is also expressed in two distinct domains, one anterior and one posterior to the *Kr* band (H. Jäckle, personal communication). These transcriptional domains are established in response to the maternally derived information encoded by the *bcd* and *osk* group genes, as is apparent from the analysis of gap gene expression in embryos lacking the *bcd* or *osk* gene products^{25,28}. In the absence of *bcd* activity, there is no zygotic expression of *hb*, whereas *Kr* is expressed more anteriorly, and in a broader domain than usual. Similarly, the absence of the posterior maternal determinants results in the extension of the *Kr* domain posteriorly. These studies suggest that the *bcd* and *osk* group genes both act to repress transcription of *Kr* in the anterior and posterior regions of the embryo, thus restricting its expression to the central region. In contrast, *bcd* acts as a positive regulator of *hb*, and by analogy it seems likely that the *osk*-dependent activity is a positive regulator of *knirps*.

Fig. 2 Schematic representation of the principal events leading to the establishment and specification of segmental primordia in the blastoderm embryo. The interactions shown have been deduced from a variety of mutant analyses, as described in the text, as well as by direct comparisons of patterns of gene expression, examples of which are shown in Fig. 4. By the time the cleavage nuclei have reached the periphery of the embryo (1.5 h after fertilization), a stable graded distribution of *bcd* protein has formed between the anterior pole and about 50% EL. At the same time it is assumed that the *osk*-dependent activity, probably the product of the *nannos(nos)* gene (R. Lehmann, personal communication), has become inhomogeneously distributed in the posterior half of the embryo. After the twelfth nuclear division, the gap genes are first zygotically transcribed (maternally deposited *hb* transcript is already present in the anterior half of the egg). Initially the transcripts are expressed in broad domains, the *hb* and *Kr* domains, overlapping considerably (unpublished observations); *hb* is activated and *Kr* is repressed by high levels of *bcd* protein. At the same time, the pair-rule genes *h* and *runt* become expressed at low levels. After a further nuclear division and before any modification of the gap gene domains, the patterns of *h* and *runt* expression begin to evolve. This process requires the activity of the gap genes. Concurrently, the gap gene domains themselves begin to narrow, a process that depends on the mutual inhibitory effects of their activities. The striped patterns of *h* and *runt* are already fully resolved by the time the gap gene domains have narrowed (see Fig. 4). The suppression of *hb* expression at the anterior pole which occurs at this stage is dependent on the activity of a third group of maternally expressed genes, the *torso* group²⁷. The homoeotic genes first become expressed in domains specified at least in part by the gap gene activities. The *h* and *runt* activities together refine the striped patterns of *ftz* and *eve*, which in turn activate the expression of *en*; in addition *ftz* enhances the expression of homoeotic genes in some of the even-numbered parasegmental primordia: the details of these latter processes are shown in Fig. 3.

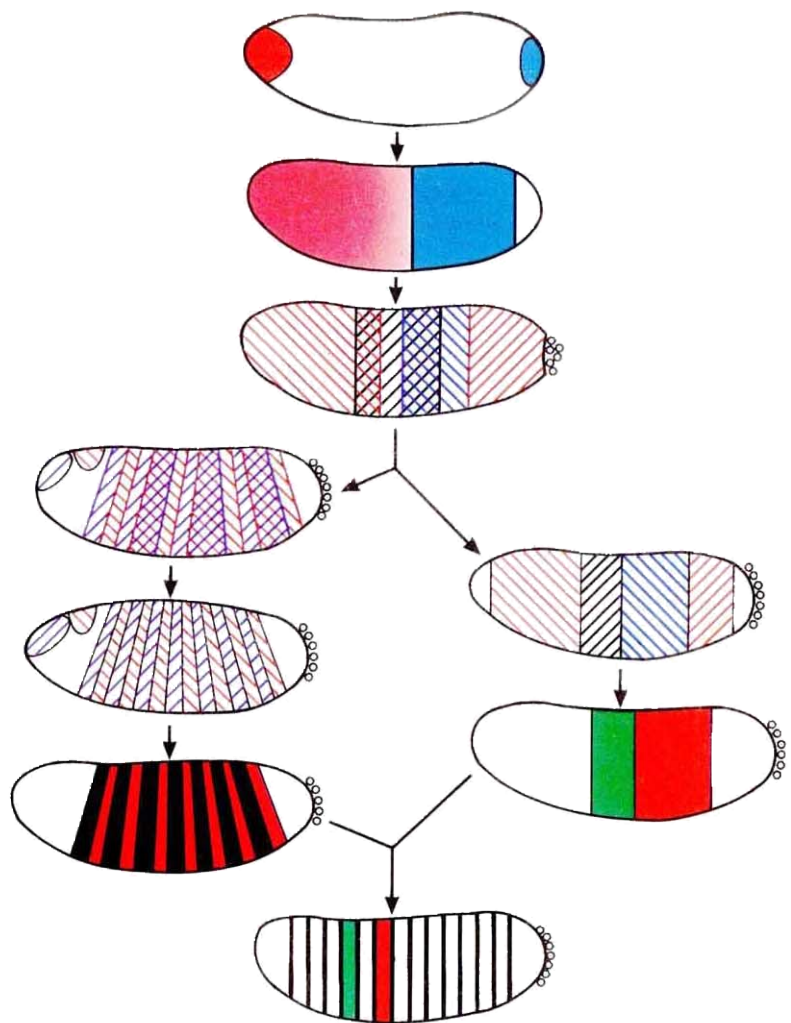
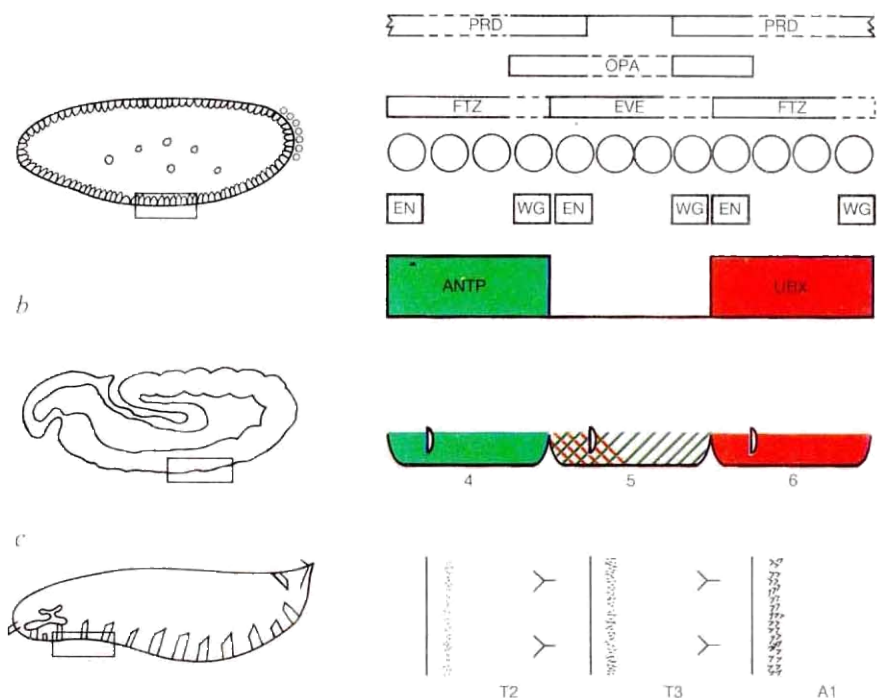


Fig. 3 Schematic representation of the *a*, establishment and specification, *b*, development and *c*, differentiation of three parasegments that contribute to the thorax and anterior abdomen of the animal. The three developmental stages, blastoderm, extended germ band and first instar larva are depicted on the left. The boxed regions are shown in detail on the right. *a*, At blastoderm, the overlapping patterns of pair-rule gene expression serve to specify the cells that will come to mark the margins of each parasegment. This is achieved by the activation of expression of the segment polarity genes *en* and *wg*. Positive regulation of *en* is by the combined activities of either *ftz* and *opa* or *eve* and *prd*. The solid lines indicate the major transcript accumulation at this stage; the dotted lines indicate earlier domains of expression, which by the end of the blastoderm stage have dissipated. *wg* appears to be activated by a release from repression as the activity of either *ftz* or *eve* decays in the most posterior row of cells of each stripe. Specification of parasegmental identity is achieved by the stimulation of homoeotic gene expression by *ftz*. Thus *Antp* transcription peaks in PS4 whereas *Ubx* peaks in PS6. Both genes are subsequently expressed at lower levels in PS5, a combined activity that uniquely characterizes this parasegment. *b*, These levels of expression are maintained as the embryo develops and are reflected in the patterns of *Antp* and *Ubx* protein distribution in the extended germ-band stage embryo⁴⁰ when the parasegments are first visible as morphological units. *c*, Each parasegment contributes to the posterior half of one body segment and the anterior half of the next. The cuticular pattern differentiated is unique to each segment and depends on the activity of *Antp*, *Ubx* and the other homoeotic genes.



This regulation is exerted at the transcriptional level, a finding consistent with the expected DNA binding activity of the *bcd* protein. Thus, high levels of *bcd* serve to activate *hb* expression, whereas low levels allow the transcription of *Kr*. Sharp boundaries are established subsequently as the transcriptional domains narrow, a process that depends on interactions between the gap genes themselves, *hb* and *Kr* mutually repressing one another, and *kni* acting as a negative regulator of *Kr*²⁹. The establishment of stable domains of gap gene expression is thus a two-step process: first a differential response to graded levels of maternal determinants, followed by mutual repression leading to the generation of stable boundaries between adjacent domains.

The information inherent in the gap gene domains is used in two distinct processes: as we shall see later, it serves to regionalize the embryo by delimiting the domains of homoeotic gene expression. In addition, it results in position-specific regulation of the pair-rule class of genes, whose periodic patterns of expression are the prelude to the metamerization of the embryo.

Initiation of periodic gene expression

There are eight members of the pair-rule class²² (an additional gene *engrailed* originally assigned to this class is now generally included in the segment-polarity class), six of which have been characterized at the molecular level (reviewed in refs 5, 6, 13 and 30; ref. 31; D. Coulter and E. Wieschaus, personal communication). A common feature of these six is their transient expression in seven or eight stripes during the cellularization of the blastoderm. Despite this basic similarity, each gene is unique in its pattern of expression. Only two, *runt* and *hairy*, are initially expressed more or less uniformly throughout the embryo; the restricted patterns of expression of these two genes begin to resolve slightly earlier than those of the other four. Moreover, in all the cases so far analysed, only *runt* and *hairy* appear to have a major function in the generation of the striped patterns of other members of the class³²⁻³⁵, and for these reasons they are referred to as the primary pair-rule genes.

By the beginning of the last interphase before cellularization, transcripts of both genes are already localized to two complementary series of seven bands that encircle the embryo; both genes are also expressed in the presumptive head region in a patch restricted to the dorsal surface of the embryo (see Figs 2 and 4). The generation of these striped patterns is a key event in the pattern-forming process, but is as yet poorly understood. What is clear is that it requires interaction both with the gap gene products and between *h* and *runt* themselves. As this process occurs before the narrowing of the gap gene domains (see Figs 2 and 4) it is likely to involve combinations of gap gene activities. The sorts of interactions taking place probably involve the local enhancement of expression of one or other gene in response to local levels of one or more gap gene products. Such interactions result in a patchy expression pattern of *h* and *runt*, which is further refined and stabilized into a regular striped pattern by their mutual repression, in a manner analogous to the stabilization of the gap gene expression domains (see Fig. 2). Evidence for this scenario comes from several different mutant analyses. First, it is clear that the generation of both the *h* and *runt* striped patterns depends on the activity of the gap genes, as aberrant patterns of expression of both genes are observed in *hb*, *Kr* and *kni* mutant embryos^{35,36}. Second, several mutations in the upstream regulatory region of *h* have been isolated, each of which eliminates the expression of the gene in specific regions of the embryo. These mutations apparently define discrete control elements which may respond to signals generated by different gap gene products³⁷. Third, the altered expression patterns of *h* and *runt* in *runt* or *h* mutant embryos, respectively, demonstrates that both genes regulate the expression of one another³⁵.

Once established, the complementary stripes of *runt* and *h* expression now serve as the driving force for the refinement of

the spatial patterns of expression of the other pair-rule genes.

Function of *hairy* and *runt* encoded proteins

Structural analyses have so far provided no clues as to the biochemical functions encoded by the *h* and *runt* genes, but it is clear from mutant studies that they act as negative regulators of the expression of the pair-rule genes *fushi tarazu* (*ftz*) and *even skipped* (*eve*) respectively. Both *ftz* and *eve* encode homoeo-domain proteins^{5,30}, and probably act as transcriptional regulators of genes of the homoeotic and segment-polarity class (see later). The interactions between the primary pair-rule genes and *ftz* and *eve* result in the generation of complementary sets of seven stripes of cells expressing one or other gene. These patterns are quite dynamic, but when they first become exactly reciprocal each stripe is on average four nuclei in width. At this stage, each *ftz* and *eve* stripe coincides with the primordium of a parasegment, metamer units that are out of register with the morphological segments but coincident with the realms of activity and function of the homoeotic genes³⁸ (Fig. 3). At the same time as the *ftz* and *eve* patterns are resolved, the pattern of expression of a third homoeobox-containing pair-rule gene, *paired* (*prd*)⁴, is also evolving. Initially, *prd* is expressed in seven broad stripes, each about six nuclei wide. This pattern is then transformed into one of fourteen stripes by the elimination of expression from the central region of each of the original stripes¹³. As with *ftz* and *eve*, the generation of this pattern depends on the activity of *h* and *runt* (M. Noll, personal communication). Thus, in a process reminiscent of the establishment of the gap gene domains, information encoded by two components is transformed into a more complex form, by the differential response of the various secondary pair-rule genes (*ftz*, *eve*, *prd*) to this information.

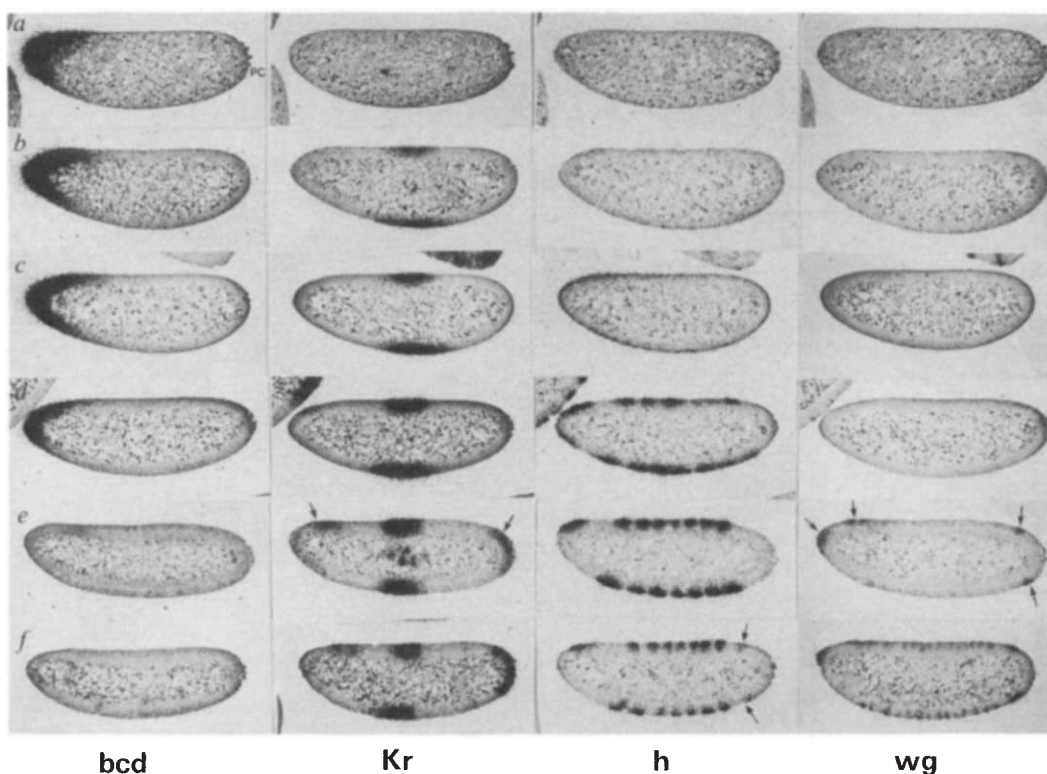
The boundaries of parasegments

The consequence of this network of interactions is that by the completion of the cellularization of the blastoderm, different cells express different combinations of pair-rule genes³⁰ (Fig. 3). There is now good evidence that these combinations of pair-rule gene activities serve to define the states of particular blastoderm cells³⁹. These cells will subsequently serve as new reference points for the elaboration of pattern within individual parasegments during the limited period of cell proliferation which follows. In particular, the pair-rule genes have been shown to define the domains of expression of the segment polarity genes *engrailed* and *wingless*³⁹, and are also responsible for the spatial regulation of transcription at the *gooseberry* locus¹³ (M. Noll, personal communication; unpublished observations).

At the end of the cellularization process and the onset of gastrulation, transcripts of both the *en* and *wg* genes accumulate in fourteen narrow stripes along the antero-posterior axis of the embryo, patterns that represent the first evidence of its segmental organization^{13,39} (Figs 2 and 4). Accumulation of *en* protein can be detected with antibodies at almost the same time, and this clearly shows that each stripe of *en* expression is on average only one cell wide⁴⁰. Each of the *en* stripes represents the anterior limit of a parasegment^{13,40}; the establishment of alternate *en* stripes requires the combined activities of different sets of pair-rule genes. In the case of the odd-numbered stripes, the *en* expressing cells are those which express both *eve* and *prd*, whereas the even-numbered stripes require the expression both *ftz* and another pair-rule gene *odd-paired*³⁹ (Fig. 3).

In contrast to this positive control of *en*, *wg* expression appears to be repressed by the *eve* and *ftz* proteins. At the end of the blastoderm stage the domains of *eve* and *ftz* narrow, so that single-cell wide stripes come to separate them³⁴, and it is these cells that initiate expression of *wg*³⁹. The net result is the generation of two sets of stripes of cells which are adjacent to one another and mark the anterior and posterior limits of each of the parasegments. At the same time, the expression of transcripts of a third segment polarity gene, *gsb*, is initiated⁴¹. These tran-

Fig. 4 Examples of the sequence of molecular patterns that generate periodicity along the antero-posterior axis of the early *Drosophila* embryo. Each row (a-f) shows four adjacent sections of the same embryo hybridized with ^{35}S -labelled probes for transcripts of *bcd*, *Kr*, *h* and *wg* respectively. The newly laid egg is endowed with maternally deposited *bcd* RNA which is localized at the anterior pole. At the end of the cleavage divisions, the nuclei migrate to the periphery of the embryo. Shortly afterwards the pole cells (PC), progenitors of the germ line, form at the posterior pole. At this stage (a) *bcd* RNA is easily detectable at the anterior pole (left), but transcripts of the three zygotically expressed genes cannot yet be detected. About ten minutes later (b) *Kr* transcripts can be seen to have



accumulated in the central third of the embryo between about 30% and 60% EL. After a further ten minutes (c), the levels of *Kr* transcript have increased; by this stage, low level *h* expression is just detectable in most of the peripherally located nuclei. After the final nuclear division before cellularization (d), *bcd* RNA is beginning to dissipate, reflecting the end of its functional period. The domains of *Kr* and the other gap genes *hb* and *kni* (not shown) are now well established (see Fig. 2) and their combined activities begin to organize the expression of *h*. Thus peaks of *h* expression appear along the antero-posterior axis in a somewhat graded fashion, a major peak at this stage coinciding with the *Kr* domain. During the next 10–15 min (e), *bcd* RNA disappears completely. Two new domains of *Kr* expression appear (arrowed), one at the anterior and one at the posterior pole of the embryo, while the central domain narrows somewhat. The modulated pattern of *h* is fully resolved into one of seven regularly spaced stripes between 70% and 20% EL and, in addition, an antero-dorsal patch of expression. At this stage, expression of the segment polarity gene *wg* is first detectable in three distinct domains: at the anterior pole; antero-dorsally, overlapping the *h* patch, and in a stripe, posterior to the last *h* stripe. *Kr* and *h* expression persists with the onset of gastrulation (f) note additional (eighth) stripe of *h* expression (arrows); by this stage, *h* activity has organized *ftz* expression (not shown), which together with *eve*, regulates expression of *wg* to be transcribed in 14 stripes, reflecting the establishment of the parasegment primordia.

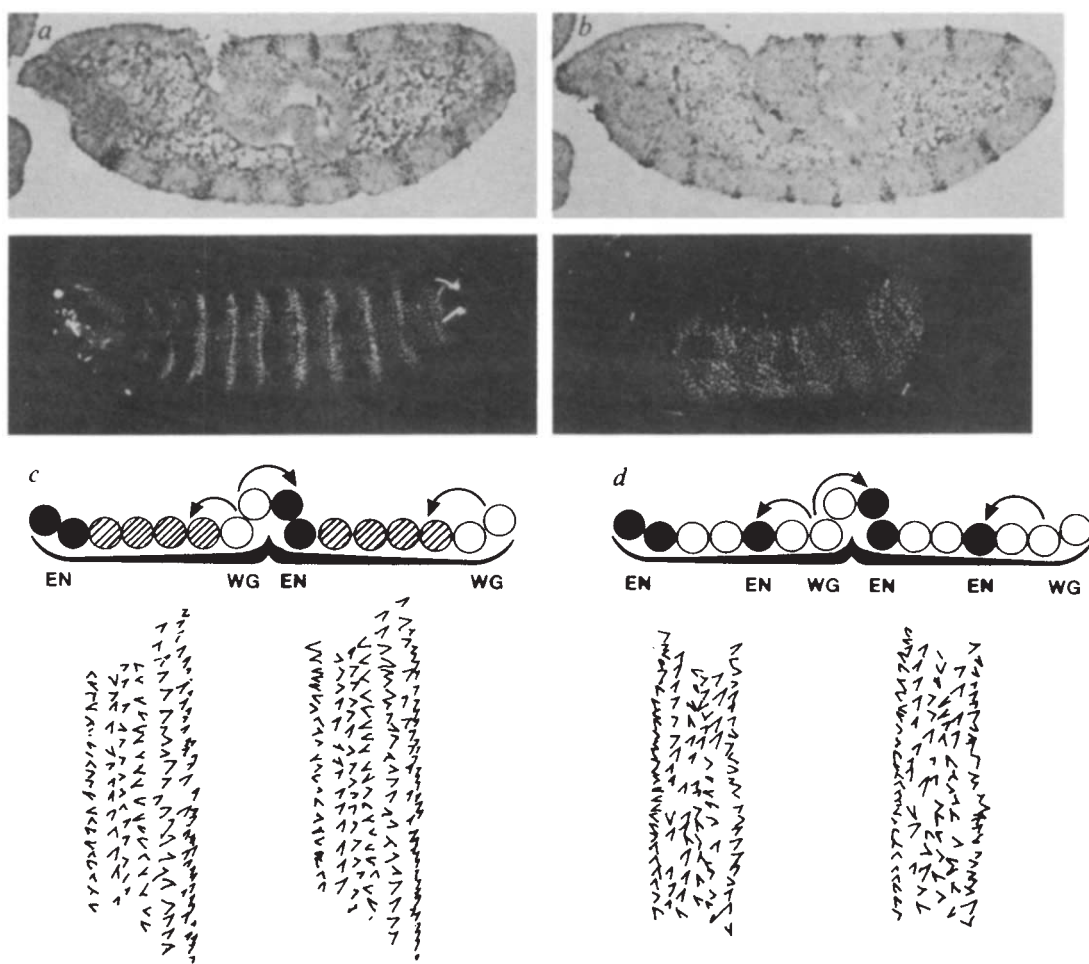
scripts accumulate in a series of 14 stripes, each two cells wide, which appear to coincide with the stripes of *prd* expression and probably therefore with both the *wg* and *en* stripes (Fig. 3). Combinatorial action of the pair-rule genes is thus crucial in specifying the cells at the anterior and posterior limits of each parasegmental primordium. These cells will subsequently serve as reference points for the specification of position in each developing parasegment.

Homoeotic genes specify parasegment character

At this stage, the embryo is not only subdivided into a series of repeating units; in addition, each parasegmental primordium is programmed to follow a unique pathway of differentiation. This specification occurs in parallel with the establishment of the *wg* and *en* stripes and depends on the selective activation of the homoeotic genes of the Antennapedia (ANT-C) and bithorax (BX-C) complexes¹³. Homoeotic genes are classically defined by their mutant phenotypes which consist of the transformation of one segment type into another (see Fig. 1). They first become transcribed just before the cellularization of the blastoderm, at about the same time as the pair-rule genes. Initially they are expressed at rather uniform levels in broad overlapping domains. The boundaries of these domains are defined both by the maternal organizing activities and by the gap genes (refs 13 and 42; V. Irish, A. Martinez-Arias and M. E. Akam, personal communication). For instance, transcripts from both promoters of the *Antennapedia* (*Antp*) gene accumulate throughout the *Kr* domain. The P1 transcript depends absolutely on *Kr* expression, whereas the P2 domain appears to be defined by a combination

of *bcd*, *osk* and *hb* expression (V. Irish, A. Martinez-Arias and M. E. Akam, personal communication). Similarly, the initial *Ubx* domain is dependent on activation by the *osk* group activity and repression by *hb* activity (ref. 13 and V. Irish, A. Martinez-Arias and M. E. Akam, personal communication). These broad regional differences are then modulated by interactions with the pair-rule prepattern to generate a series of unique parasegmental states. Expression of *Scr*, *Antp* and *Ubx* is specifically elevated in the primordia of parasegments 2, 4 and 6, respectively, by the activity of *ftz*⁴³ (Fig. 3). These enhanced levels of transcription are subsequently maintained throughout embryogenesis and are reflected in the mutant phenotypes of each gene. Thus, *Scr* mutant embryos exhibit a transformation of the labial segment, part of which derives from PS2, whereas *Antp* mutants show a transformation of the second thoracic segment, a derivative of PS4. Similarly, *Ubx* mutant larvae have their first abdominal segment (derived from PS6) transformed to a second thoracic segment. In addition, *Ubx* mutants also show subtle effects on the differentiation of the other abdominal segments, reflecting the lower level expression of *Ubx* in PS7–13 which persists throughout embryogenesis¹³. This complex and evolving pattern of expression is in fact a general feature of homoeotic genes; indeed the expression and mutant effects of *Scr* and *Antp* extend beyond the PS2 and PS4 domains described above¹³. Although the ontogeny of the homoeotic gene domains is still incompletely understood, it is clear that it is dependent in part on interactions between the genes themselves. In general, the more posteriorly expressed genes (such as *abd A* and *Abd B* of the BX-C) repress the expression of the more anteriorly

Fig. 5 Induction and maintenance of cell states along the antero-posterior axis. *a* and *b*, The patterns of expression and mutant phenotypes of the segment polarity genes *gsb* and *wg* respectively. Although *wg* is expressed in a smaller domain than *gsb* in each parasegment, its absence has a much more drastic effect on the final pattern. This presumably reflects the different functions of the two genes. Whereas *gsb* is probably required autonomously for the specification of just those cells in which it is expressed, the activity of *wg* extends beyond its transcriptional domain. The possible basis of the *wg* function is shown schematically in *c* and *d*. *c*, Representation of the ectoderm of two adjacent parasegments in a normal extended germ-band embryo and below, the ventral cuticular pattern that they differentiate. Cells at the anterior (left) of each parasegment express *en* (shaded), cells at the posterior express *wg* (open). The maintenance of *en* expression depends on the continuous expression of *wg*, as indicated by the arrows. Cells anterior to those expressing *wg* may also receive the *wg* signal, but respond differently as a result of the expression of *ptc* (hatched). *d*, In the absence of *ptc* these cells also express *en*, resulting in an altered pattern of cell states and ultimately in an altered pattern in the larval cuticle, as indicated.



expressed genes (such as *Ubx* and *Antp*)¹³. As well as generating differences between parasegments, homoeotic genes are also expressed differentially between germ layers and within each metamere. This differential expression serves to modulate the basic pattern that underlies the organization of each segment. To see how this basic pattern is generated, we must return to a consideration of the segment polarity genes.

Elaboration of pattern in parasegments

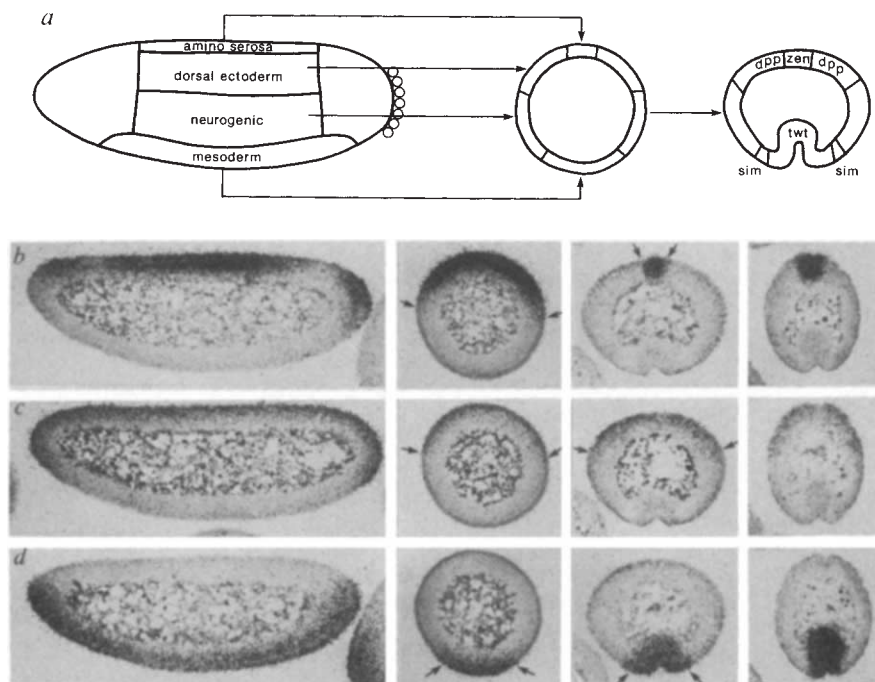
In the preceding sections we have seen how the cells at the boundaries of each parasegment are specified, a specification reflected by their expressing particular segment polarity genes. Although it is possible that the intervening blastoderm cells likewise become determined in response to pair-rule gene signals⁴⁴, a more likely hypothesis is that they remain unspecified at this stage. Specification of cells within each parasegment would then occur as the cells divide in response to signals mediated by cell-cell interactions. In line with this model is the finding that another segment-polarity gene, *patched* (*ptc*), is initially expressed uniformly in the cellular blastoderm and early gastrulating embryo, and only becomes spatially restricted in its expression at the extended germ band stage (I. Guerrero, A. Hidalgo, A. Taylor, R. Whittle and P.W.I., unpublished results). The implication that cells are responding to signals during this stage is supported by analysis of the *wingless* gene. Although *wg* is only transcribed in less than one quarter of the cells of each parasegment^{45,46}, its absence results in the elimination of about three quarters of the pattern elements from each segment (Fig. 5). This suggests that the role of *wg* extends beyond its transcriptional domain and into neighbouring cells, a conclusion

supported by several lines of evidence. First, mosaic analysis of *wg* mutants has shown that the effect of the mutation is not cell-autonomous⁴⁷. Second, the continuous expression of *en* by cells in the anterior part of each parasegment requires the expression of *wg* in cells adjacent to them^{46,48}. Third, *wg* encodes a protein that exhibits significant homology with the mouse *int-1* gene product, a putative secretory protein⁴⁹. Taken together, this evidence suggests that cells expressing *wg* are capable of signalling to their neighbours. The response to this signal may depend on the competence of the cells receiving it; thus, whereas cells posterior to those expressing *wg* are already competent to express *en*, those anterior to the *wg* domain acquire a different fate (see Fig. 5). This differential response may depend on the activity of the *ptc* gene. Indeed, the pattern of expression of *ptc* is consistent with its endowing cells with the capacity to respond differentially to a *wg*-encoded signal. Moreover, in the absence of *ptc*, the *wg* domain expands and cells on both sides of it now express *en*^{46,48}.

The implication of these studies is that the final fine-grained pattern along the antero-posterior axis is generated first by establishing a series of reference points (the *en*- *wg*- and *gsb*-expressing cells), and then elaborating information between these reference points by the interaction between cells. Thus, for instance, as cells divide they would assume states dependent on their apposition to *wg*-expressing (or *en*-expressing) cells. Further divisions would allow the intercalation of additional cell states⁴⁶. Such a process requires the existence of signalling molecules and receptors capable of transducing and receiving information between cells.

An analogous process has been shown to underlie the genera-

Fig. 6 *a*, Schematic representation of the specification of fate along the dorso-ventral axis of the embryo (adapted from ref. 52). The regions of the blastoderm embryo having different fates are shown in longitudinal (left) and transverse (centre) section. These different fates are reflected in the spatially restricted patterns of expression of zygotically required genes (right). *b*, *c* and *d*, Adjacent sections of successively older embryos which have been hybridized with probes for the transcripts of *zen*, *dpp* and *twi* respectively. The embryo on the left is oriented longitudinally, anterior to the left, the remainder transversely. Dorsal is uppermost in all cases. At the end of the blastoderm stage (left), transcripts of each gene accumulate either dorsally (*zen* and *dpp*) or ventrally (*twi*). The initial domains of *zen* and *dpp* are almost coincident. But as the embryo begins to gastrulate (centre right), the *zen* domain narrows considerably, whereas the *dpp* domain remains constant. Concurrently, *twi* expression is initiated in all the cells that will subsequently invaginate during gastrulation (centre right) to form the mesoderm.



tion of pattern in the adult *Drosophila* eye⁵⁰. Each individual unit, or ommatidium, of the compound eye is composed of a precise array of cells of varying type. The generation of this unit begins with the specification of a core of cells; subsequently, additional cells are recruited and specified in accordance with the contacts they make with the core cells. Recently the *sevenless* gene (which when mutated results in the transformation of a specific photoreceptor cell, R7) has been shown to encode a transmembrane protein⁵¹. This has both an intracellular tyrosine kinase domain and an external domain with homology to the repeated motif of the mammalian epidermal growth factor (EGF). The *sevenless* product therefore appears to be a receptor responsible for transducing the inductive signal that specifies the fate of the R7 cell. The *ptc* gene might similarly encode a receptor protein that determines the response of particular cells to inductive signals. Certainly we can expect at least some of the segment polarity genes to encode proteins involved in signal transduction.

Specification of dorso-ventral pattern

All of the foregoing interactions can, of course, only provide cells with information about their position along the antero-posterior axis. For cells to know whether they should differentiate into epidermis, gut, neurons or muscles, they must also be provided with information as to their position along the dorso-ventral axis. Concurrent with the segmental subdivision of the embryo, a separate but analogous chain of genetic interactions directs its partitioning along this axis. The details of these interactions are at present less well characterized than those between the segmentation genes, but it is clear that they too involve a transfer of information between maternally and zygotically expressed genes.

Genes defining dorso-ventral polarity

The maternally derived determinants of dorso-ventral polarity are encoded by a number of genes, most of which were identified as maternal-effect mutations producing dorsalized embryos⁵². Females lacking any one of these genes lay eggs that develop into embryos devoid of any lateral or ventral derivatives (that is, mesoderm and ventral cuticle). Three of these genes have now been cloned and all exhibit sequence homologies with vertebrate genes⁵³.

In contrast to *bed*, there is no evidence for a localization of the transcripts of any of these genes in the egg, suggesting that dorso-ventral polarity is generated by a mechanism different from that organizing antero-posterior polarity. Much of what is known about their function has been gleaned from experiments involving the rescue of mutant embryos by transplantation of wild-type cytoplasm or RNA⁵². Mutant *snake* embryos can be rescued by cytoplasm from any region of a wild-type embryo, irrespective of the site of injection. This implies that *snake* activity is not in itself sufficient for generating asymmetry. The *snake* protein shows significant homology with human serine protease, suggesting a role for it in the proteolytic cleavage of other spatially localized products⁵⁴.

A better candidate for a gene encoding such a product is *dorsal* (*dl*). Although *dl* transcript is homogeneously distributed in the early embryo⁵⁵, the rescuing activity becomes enriched ventrally towards the end of the blastoderm stage, suggesting a heterogeneity in the distribution of *dl* protein. Mutant *dl* embryos can only be rescued if the transplanted material is injected ventrally⁵², implying that the egg is endowed with information along its dorso-ventral axis independently of *dl* activity. Such information appears to be generated by the product of the *Toll* gene. Rescue of embryos lacking *Toll* activity can be effected with cytoplasm taken from any region of a wild-type egg, but, in contrast to all other mutants of this class, the site of injection defines the position at which the ventral-most structures will form⁵². Molecular analysis of *Toll* has revealed that, like *dl*, its transcript is uniformly distributed in the early embryo⁵⁶. This finding is in line with the fact that poly A⁺ RNA from eggs can efficiently mimic the non-localized rescuing activity of wild-type cytoplasm⁵². Together these results suggest that it is the *Toll* protein that becomes localized or modified in the presumptive ventral region of the embryo, a process that apparently depends on *Toll* activity itself⁵³. Sequence analysis of *Toll* has shown that it encodes an integral membrane protein⁵⁶. Whether this acts in turn to localize the *dl* protein remains to be seen.

Subdivision along the dorso-ventral axis

Whatever the mechanism, the maternal dorsalizing genes generate information which is then interpreted, like that along the antero-posterior axis, by the zygotic genome, resulting in the

spatially restricted expression of genes around the circumference of the embryo (see Fig. 6).

These genes have been identified both on the basis of their mutant phenotypes and by their sequence homologies. Embryos mutant for the *zen* and *dpp*^{Hind} genes lack normal dorsal derivatives and exhibit a 'ventralized' phenotype^{57,58}. By contrast, absence of the *twist* and *snail* genes results in the elimination of the mesoderm⁵⁹, whereas *single-minded* mutants eliminate both neuronal and non-neuronal derivatives of the ventral-most ectoderm^{60,61}. As with the pair-rule genes, the domains of expression of some of these genes are initially quite broad and overlapping, but become refined into narrower stripes (see Fig. 6). In the case of *zen*, this refinement is completed by early gastrulation⁶² and may well be accomplished by the direct interaction between *zen* and the maternally encoded gene products⁶³. In other cases, the refinement is more gradual and occurs later than that of the pair-rule genes⁶⁵, after the embryo has cellularized. Thus it could depend on interactions between cells. Interestingly, the *dpp*^{Hind} product shows significant homology with the mammalian growth factor TGF- β (ref. 65), and more laterally expressed gene was identified by its homology with EGF⁶⁶. Such proteins would be expected to mediate cell-cell signalling and could therefore contribute to the refinement of their own patterns of expression.

This gradual refinement of the patterns of the more dorsally expressed genes during gastrulation is consistent with the finding that the cells of the dorsal ectoderm remain pluripotent for some time after the blastoderm stage¹⁰. But, moving more ventrally around the embryo, the potential of the cells becomes progressively more restricted. This earlier commitment is paralleled by the finding that the patterns of expression of the ventrally restricted zygotic genes are much more precisely defined from the onset. Thus *twist* is expressed in just those blastoderm cells that will invaginate into the ventral furrow and form the mesoderm⁶⁷, whereas *single minded* is expressed in two narrow stripes of cells either side of and adjacent to the *twist* domain^{60,61} (see Fig. 6). Both *sim* and *twist* encode nuclear proteins^{61,67}, consistent with their being directly responsible for the determination of the cells in which they are expressed. How the precise spatial control of their expression is achieved is unclear. It could reflect a direct response of these genes to the maternally encoded information⁶⁷; alternatively, other zygotic genes, as yet uncharacterized, may intervene in refining this information in a manner similar to the operation of the gap genes along the antero-posterior axis. Interestingly, the *snail* gene shares structural homology with the gap genes in that it contains a series of DNA-binding finger domains⁶⁸. Whether or not *sna* regulates the expression of *twist* or *sim* remains to be seen.

This sequence of events suggests an interesting parallel between specification of cell fate along the antero-posterior axis within each parasegment and along the dorso-ventral axis. Just as the anterior-most cell of each parasegment seems to be rigidly determined at blastoderm (by activation of the nuclear protein *en*), so are the ventral-most cells (by activation of *twist*). In contrast, the specification of the more posterior, or more dorsal cells, occurs progressively under the influence of genes, some of which encode growth-factor related proteins.

Neurogenesis

The development of the nervous system provides a good example of the way in which cells interpret and respond to the orthogonal array of information established at blastoderm. The ventral nervous system derives from cells of the ventral ectoderm, that is from the cells that at gastrulation lie roughly between the ventral midline (namely adjacent to the mesoderm) and the lateral midline (Fig. 6). The allocation of cells to the neural pathway of development therefore depends on information along the dorso-ventral axis—in mutant dorsalized embryos the number of cells entering this pathway is considerably reduced⁶⁹. Of zygotically active genes that register the dorso-ventral infor-

mation, *sim* has been shown to be involved in specifying part of the neurogenic domain. Thus in the absence of *sim*⁺ function, specific neurons fail to differentiate. But *sim* mutants also lack non-neuronal cells, indicating that this gene provides cells only with information about their position along the dorso-ventral axis. This endows cells with the potential to follow the neurogenic pathway, and a second mechanism operates to limit the number of cells that enter it. This mechanism depends on the lateral inhibition of neighbouring cells by nascent neuroblasts, so that inhibited cells accordingly follow the dermal differentiation pathway⁷⁰. A number of so-called neurogenic genes have been implicated in this mechanism, absence of any of which results in the hypertrophy of the nervous system^{69,70}. Several of these genes have molecular structures consistent with their involvement in intercellular signalling. For instance *Notch* encodes a transmembrane protein with an EGF-related extracellular domain⁷⁰. This gene interacts with the *Enhancer of split* (*E(spl)*) locus, the product of which is homologous to the β -subunit of mammalian G-proteins (D. Hartley, A. Priess and S. Artavanis-Tsakonas, personal communication).

Once a neuroblast has segregated from the ectoderm, it will give rise to a specific neural lineage dependent on its position along the antero-posterior axis⁷¹. This information is presumably generated by the segmentation genes and is interpreted by other genes specifically required for neuronal differentiation. Some of these are encoded by a complex of transcription units known as the achaete-scute complex⁷². These are first transcribed at the end of the blastoderm stage in striped patterns reminiscent of *wg* or *en*, which rapidly become restricted to the ventral ectoderm⁷². Expression of these transcripts appears to render cells competent to enter the neuronal pathway, and different transcripts presumably determine different lineages.

Redeployment of segmentation genes

The molecular analysis of the segmentation genes has revealed a further unexpected function for them in neurogenesis, hitherto undetected by their genetic analysis. First *ftz*, and subsequently *eve* and *Kr*, have all been shown to undergo a second distinct phase of expression in specific tissues, primarily in developing neuronal cells, in the case of *eve* and *ftz*^{71,73,74}. In addition, and perhaps less unexpectedly, the segment polarity genes *en* and *wg* are also expressed in the nervous system. All of these genes exhibit highly specific patterns of expression in distinct but overlapping sets of neuroblasts or neurons. In contrast to the earlier blastoderm expression patterns of *Kr*, *eve* and *ftz*, they are all expressed in a segmentally reiterated manner. Analysis of the function of this second phase of expression is so far limited, but the results to date suggest that the genes are involved in specifying neuronal fate^{73,74}. By specifically eliminating the neuronal expression of *ftz*, it has been shown that identified neurons that normally express the gene display altered patterns of axonal projection⁷⁴. Intriguingly, absence of *ftz* expression from some neurons results also in the elimination of *eve* expression, implying a regulatory relationship between these two genes which is the converse of that at blastoderm.

Perspectives

The emerging picture is one of two distinct phases in the development of the embryo. The first begins at fertilization and is more or less completed by the time cells are forming at the end of the blastoderm stage. Many of the genes involved in this process appear to encode transcription factors/nuclear proteins suggesting that the process is accomplished by a transcriptional regulation cascade. The successive subdivision of the embryo into smaller and smaller domains seems to be accomplished by the differential and combinatorial action of transcription factors, and is initiated by the activities of a limited number of localized determinants. The next phase that begins after the cellular blastoderm consists of the elaboration of the information provided by the reference points along the orthogonal axes. This

process requires the communication of information between cells, and accordingly depends on molecules capable of inter-cellular signalling.

Despite the unusual nature of early *Drosophila* embryogenesis, with the initial patterning events occurring in a syncytium, many of the principles that are being established from its study are likely to have wide implications. The localization of transcripts, whether to specific regions of the embryo or to particular compartments of the cell, as clearly demonstrated for many of the segmentation genes, has its counterpart in the *Xenopus veg-1* gene⁷⁵, and is certain to be a fundamentally important device in the development of most organisms. The potential for investigating the control of such localization in *Drosophila* is immense. Already mutations have been identified in genes required for the localization of *bcd*, and the ability to reintroduce modified genes will permit a precise dissection of the RNA/cytoskeletal interactions implied by this localization.

Many of the nuclear protein-encoding genes are characterized by one of two conserved motifs—the homoeobox and the nucleic acid-binding fingers. The implication of these proteins in transcriptional regulation comes from the presence of similar motifs in known transcriptional regulators—the yeast *mat* genes in the case of the homoeobox⁵, and *Xenopus TF11A* and human *SP1* in the case of the finger motif⁷⁶. Intriguingly, not only is the structure conserved between yeast and *Drosophila*, but both types of motif are associated with genes controlling cell fate. Switching between mating types in *S. cerevisiae* involves the activation of different sets of genes that are under the control of the *mat a* and α transcriptional regulators, the DNA-binding domains of which are homologous to the homoeobox. Regulation of these genes in turn depends indirectly on the function of a finger-motif protein encoded by the *swi-5* gene. Thus, regulatory interactions between a finger-motif protein and homoeo-domain proteins serve to determine cell fate in yeast⁷⁷. In an analogous regulatory relationship, finger-motif proteins (*Kr*, *kni* and *hb*) and homoeo-domain proteins (*ftz* and *Ubx*) determine cell fate in *Drosophila*. This conservation of function and of sequence encourages the view that the motifs will identify functionally similar genes in a wide variety of organisms. The recent identification of a finger-motif in the human testis-determining factor⁷⁸ suggests a similar function for at least some vertebrate finger proteins.

Many mouse homoeobox-containing genes have now been identified and their spatial and temporal patterns of expression analysed in detail. Some of these are deployed in a strikingly similar manner to their *Drosophila* counterparts, being expressed in restricted but overlapping domains along the antero-posterior axis of the embryo⁷⁹. Thus it seems quite plausible that these genes are involved in some aspect of regionalization in a manner analogous to the ANT-C and BX-C genes.

The challenge now is to discover the genes downstream of these putative transcriptional regulators, for which again *Drosophila* is attractive as an experimental organism. Although such genes have yet to be identified by direct binding assays, other approaches have successfully yielded mutations in what might properly be regarded as 'morphogenetic' genes. The product of the *PS2* gene⁸⁰ was originally identified using antibodies directed against spatially restricted cell-surface antigens. Cloning and sequencing of the gene have shown it to be related to the integrin family of fibronectin receptors⁸⁰. The gene is first expressed during early embryogenesis in the cells of the prospective mesoderm and is subsequently required for a number of morphogenetic processes, including muscle attachment. The remarkable correlation between its pattern of transcription with the expression of *twist* makes *PS2* a good candidate for a gene regulated by the *twist* nuclear protein. Indeed, *PS2* expression is abolished in *twist* mutants (M. Wilcox, personal communication). A similar immunological approach has led to the identification of the *fasciclin* genes which have structural similarities to the vertebrate cell adhesion molecules⁸¹. Their products are

selectively expressed on axon growth cones of different neurons and may therefore be regulated by the various homoeobox genes expressed in different neural lineages.

Perhaps even more significant is the growing number of *Drosophila* genes having sequence homology with growth factors and their receptors. The role of the TGF- β related gene *dpp*^{Hind} in determining the fate of dorsal ectoderm in *Drosophila* is paralleled by recent findings in *Xenopus*^{82,83}. Here it has been shown that TGF- β and related proteins, as well as fibroblast growth factor (FGF), can induce animal pole ectoderm to become mesoderm. Such early functions for members of growth factor families are likely to be widespread, as indicated by other recent studies in mouse. Analysis of the expression of the FGF-related *int-2* gene by *in situ* hybridization has revealed that it is expressed as early as at 7.5 days of development in mesodermal cells⁸⁴. Similarly *int-1*, the murine homologue of the *Drosophila* *wg* gene, is also expressed in a spatially restricted manner in early mouse embryos, in this case in cells of the developing nervous system⁸⁵. The elucidation of the molecular and cellular mechanisms of the activities of such genes in *Drosophila*, facilitated by the ease of its genetic analysis, is thus likely to yield insights into the principles of early inductive events and cellular interactions in a wide range of organisms.

I thank Ken Howard, Julian Lewis, Paul Martin, Jay Mitten-thal and Alfonso Martinez-Arias for discussion; Kate Hooper and David Ish Horowitz for unpublished observations of the *hairy* protein; Thomas Berleth, Fabienne Perrin-Schmitt, Chris Rushlow and Daniel St Johnston for DNA probes, and Pam Hickey for preparation of the manuscript.

1. Wilson, E. B. *The Cell in Development and Heredity* 3rd edn (Macmillan, New York, 1925).
2. Lewis, E. B. *Cold Spring Harb. Symp. quant. Biol.* **16**, 159–174 (1951).
3. Lewis, E. B. *Nature* **276**, 565–570 (1978).
4. Frigerio, G., Burri, M., Bopp, D., Baumgartner, S. & Noll, M. *Cell* **47**, 735–746 (1986).
5. Gehring, W. J. *Science* **236**, 1245–1252 (1987).
6. Gehring, W. J. & Hiromi, Y. *A. Rev. Genet.* **20**, 147–173 (1986).
7. Sander, K. *Adv. Insect Physiol.* **12**, 125–238 (1976).
8. Foe, V. E. & Alberts, B. M. *J. cell Sci.* **61**, 31–70 (1983).
9. Campos-Ortega, J. A. & Hartenstein, V. *The Embryonic Development of Drosophila melanogaster* (Springer, Berlin, 1985).
10. Technau, G. M. *Development* **100**, 1–12 (1987).
11. Wieschaus, E. & Gehring, W. *Dev. Biol.* **50**, 249–263 (1976).
12. Simcox, A. A. & Sang, J. H. *Dev. Biol.* **97**, 212–221 (1983).
13. Akam, M. E. *Development* **101**, 1–22 (1987).
14. Schubiger, G. & Newman, S. M. *Am. Zool.* **22**, 47–56 (1982).
15. Nüsslein-Volhard, C., Frohnhofer, H. G. & Lehmann, R. *Science* **238**, 1675–1681 (1987).
16. Frohnhofer, H. G. & Nüsslein-Volhard, C. *Nature* **324**, 120–125 (1986).
17. Berleth, T. et al. *EMBO J.* **7**, 1749–1756 (1988).
18. Lehmann, R. & Nüsslein-Volhard, C. *Cell* **47**, 141–152 (1986).
19. Driever, W. & Nüsslein-Volhard, C. *Cell* **54**, 89–98 (1988).
20. Driever, W. & Nüsslein-Volhard, C. *Cell* **54**, 95–104 (1988).
21. Lehmann, R. & Nüsslein-Volhard, C. *Nature* **329**, 167–170 (1987).
22. Nüsslein-Volhard, C. & Wieschaus, E. *Nature* **287**, 795–801 (1980).
23. Rosenberg, U. B. et al. *Nature* **319**, 336–339 (1986).
24. Tautz, D. et al. *Nature* **327**, 383–389 (1987).
25. Tautz, D. *Nature* **332**, 281–284 (1988).
26. Knipple, D. et al. *Nature* **317**, 40–44 (1985).
27. Gaul, U., Seifert, E., Schuh, R. & Jäckle, H. *Cell* **50**, 639–647 (1987).
28. Gaul, U. & Jäckle, H. *Cell* **51**, 549–555 (1987).
29. Jäckle, H., Tautz, D., Schuh, R., Seifert, E. & Lehmann, R. *Nature* **324**, 668–670 (1986).
30. Scott, M. P. & Carroll, S. B. *Cell* **51**, 689–698 (1987).
31. Gergen, J. P. & Butler, B. A. *Genes Dev.* (in the press).
32. Howard, K. R. & Ingham, P. W. *Cell* **44**, 949–957 (1986).
33. Carroll, S. B. & Scott, M. P. *Cell* **45**, 113–126 (1986).
34. Frasch, M. & Levine, M. *Genes Dev.* **1**, 981–995 (1987).
35. Ingham, P. W. & Gergen, J. P. *Development* (Suppl.) (in the press).
36. Ingham, P. W., Ish-Horowitz, D. & Howard, K. R. *EMBO J.* **5**, 1659–1665 (1986).
37. Howard, K. R., Ingham, P. W. & Rushlow, C. A. *Genes Dev.* (in the press).
38. Martinez-Arias, A. & Lawrence, P. A. *Nature* **313**, 639–642 (1985).
39. Ingham, P. W., Baker, N. E. & Martinez-Arias, A. *Nature* **331**, 73–75 (1988).
40. Carroll, S. B. et al. *Genes Dev.* **2**, 350–360 (1988).
41. Baumgartner, S., Bopp, D., Burri, M. & Noll, M. *Genes Dev.* **1**, 1247–1267 (1987).
42. Harding, R. & Levine, M. *EMBO J.* **7**, 205–214 (1988).
43. Ingham, P. W. & Martinez-Arias, A. *Nature* **324**, 592–597 (1986).
44. Gergen, J. P., Coulter, D. & Wieschaus, E. *Symp. Soc. dev. Biol.* **44**, 195–220 (1986).
45. Baker, N. E. *EMBO J.* **6**, 1765–1773 (1987).
46. Martinez-Arias, A., Baker, N. E. & Ingham, P. W. *Development* **103**, 157–170 (1988).
47. Morata, G. & Lawrence, P. A. *Dev. Biol.* **56**, 227–240 (1977).
48. DiNardo, S., Sher, E., Heemskerk-Jongens, J., Kassis, J. A. & O'Farrell, P. H. *Nature* **332**, 604–609 (1988).
49. Rijsewijk, F. et al. *Cell* **50**, 649–657 (1987).
50. Tomlinson, A. & Ready, D. F. *Dev. Biol.* **120**, 366–376 (1987).
51. Hafen, E., Basler, K., Edstroem, J. E. & Rubin, G. M. *Science* **236**, 55–63 (1988).
52. Anderson, K. V. *Trends Genet.* **3**, 91–97 (1987).
53. Levine, M. *Cell* **52**, 785–786 (1988).
54. DeLotto, R. & Spierer, P. *Nature* **323**, 688–692 (1986).

55. Steward, R. *et al.* *Cold Spring Harbor Symp. quant. Biol.* **50**, 223-228 (1985).
56. Hashimoto, C., Hudson, K. L. & Anderson, K. V. *Cell* **52**, 269-279 (1988).
57. Wakimoto, B. T., Turner, F. R. & Kaufman, T. C. *Dev. Biol.* **102**, 147-172 (1984).
58. Irish, V. F. & Gelbart, W. M. *Genes Dev.* **1**, 868-879 (1987).
59. Simpson, P. *Genetics* **105**, 615-632 (1983).
60. Thomas, J. B., Crews, S. T. & Goodman, C. S. *Cell* **52**, 133-141 (1988).
61. Crews, S. T., Thomas, J. B. & Goodman, C. S. *Cell* **52**, 143-151 (1988).
62. Doyle, H. J., Harding, K., Hoey, T. & Levine, M. *Nature* **323**, 76-79 (1986).
63. Rushlow, C., Frasch, M., Doyle, H. & Levine, M. *Nature* **330**, 583-586 (1987).
64. St Johnston, R. D. & Gelbart, W. M. *EMBO J.* **6**, 2785-2791 (1987).
65. Padgett, R. W., St Johnston, R. D. & Gelbart, W. M. *Nature* **325**, 81-84 (1987).
66. Knust, E. *et al.* *EMBO J.* **6**, 761-766 (1987).
67. Thisse, B., Stoetzel, C., Gorostiza-Thisse, C. & Perrin-Schmitt, F. *EMBO J.* **7**, 2175-2184 (1988).
68. Boulay, J. L., Dennefeld, C. & Alberga, A. *Nature* **330**, 395-38 (1987).
69. Campos-Ortega, J. A. *Trends Neurosci.* **8**, 245-250 (1985).
70. Artavanis-Tsakonas, S. *Trends Genet.* **4**, 95-100 (1988).
71. Doe, C. Q. & Scott, M. P. *Trends Neurosci.* **3**, 101-106 (1988).
72. Cabrera, C. V., Martinez-Arias, A. & Bate, M. *Cell* **50**, 425-433 (1987).
73. Doe, C. Q., Smouse, D. & Goodman, C. S. *Nature* **33**, 376-378 (1988).
74. Doe, C. Q., Hiromi, Y., Gehring, W. J. & Goodman, C. S. *Science* **239**, 170-175 (1988).
75. Weeks, D. C. & Melton, D. A. *Cell* **51**, 861-867 (1987).
76. Evans, R. M. & Hollenberg, S. M. *Cell* **52**, 1-3 (1988).
77. Nasmyth, K. & Shore, D. *Science* **237**, 1163-1170 (1987).
78. Page, D. C. *et al.* *Cell* **51**, 1091-1104 (1987).
79. Gaunt, S. J., Sharpe, P. T. & Duboule, D. *Development (Suppl.)* (in the press).
80. Bogaert, T., Brown, N. & Wilcox, M. *Cell* **51**, 929-940 (1987).
81. Snow, P. M. *et al.* *Proc. natn. Acad. Sci. U.S.A.* (in the press).
82. Kimelmann, D. & Kirschner, M. *Cell* **51**, 869-877 (1987).
83. Slack, J. M. W., Darlington, B. G., Heath, J. K. & Godsave, S. *Nature* **326**, 197-200 (1987).
84. Wilkinson, D. G., Peters, G., Dickson, C. & MacMahon, A. P. *EMBO J.* **6**, 691-695 (1988).
85. Wilkinson, D. G., Bailes, J. A. & McMahon, A. P. *Cell* **50**, 79-88 (1987).

ARTICLES

Implications for continental structure and evolution from seismic anisotropy

Paul G. Silver & W. Winston Chan*

Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road NW, Washington DC 20015, USA

*Teledyne Geotech Alexandria Laboratories, 314 Montgomery Street, Alexandria, Virginia 22314, USA

Evidence is presented for azimuthal anisotropy in the subcontinental upper mantle from nine stations in North America and Europe. For the stations in the Canadian Shield, this is due to 2.5-2.7 Gyr-old 'fossil' anisotropy localized in the top 200 km of the upper mantle and produced by the preferred orientation of mantle minerals. This mantle fabric, apparently caused by a series of large-scale deformational episodes at the end of the Archean, has moved coherently with the surface ever since.

SEISMIC anisotropy is becoming increasingly important in our understanding of the Earth's crust and mantle. Its close relationship to the strain-induced preferred orientation of highly anisotropic crystals (such as olivine) provides a method of studying motions in the mantle associated with plate tectonics, as well as the internal deformation of the plates. Anisotropy in the oceanic upper mantle reveals a remarkably simple picture. Studies of P waves that travel just below the crust (P_n) invariably exhibit a fast direction that is parallel to the fracture zones¹⁻³, strongly suggesting that the anisotropy is due to fossil fabric formed at the spreading ridge. Azimuthal anisotropy, inferred from long period surface waves^{4,5} which sample below the oceanic plates, indicates a fast direction that is consistent with the flow directions inferred from simple kinematic models of mantle convection.

Anisotropy in the continents is less well documented and more difficult to study. Few P_n studies have been done⁶⁻⁸. Some shear-wave splitting measurements have been made for local events in the upper crust in Turkey^{9,10} and other regions¹¹, although no general characterization of continental anisotropy has emerged. Azimuthally dependent P-wave station residuals (second azimuthal term) have been estimated for most continental stations¹², but these are relatively indirect measures of anisotropy. The lack of information on continental anisotropy is largely due to the complicated pattern of deformation in the continents when compared to the oceans. The continents not only have had a longer history, but the length scales of coherent deformation are much smaller as the surface geology quickly reveals.

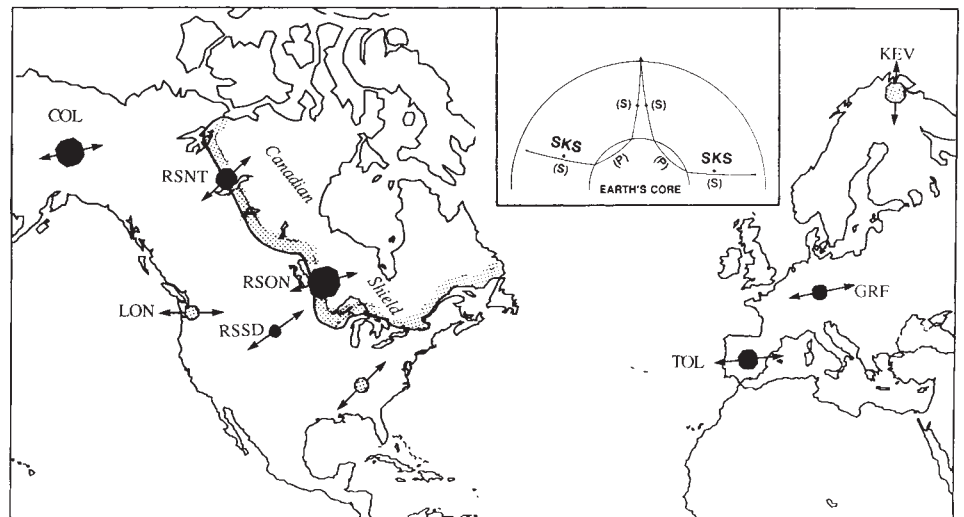
A manifestation of seismic anisotropy that is ideally suited for such an environment is shear-wave splitting¹³, where an incident shear-wave is polarized into two orthogonal directions

travelling at different velocities. The polarization directions and the delay time between the two arrivals provide constraints on the orientation of the principal axes of strain, in addition to the depth extent of the anisotropy and its magnitude. This phenomenon is analogous to birefringence in minerals. The near-vertical paths through the mantle, which are the most advantageous for observing shear-wave splitting, provide excellent lateral resolution: anisotropy can thus provide information on strain at depth that can be compared to past or present surface deformation.

The shear waves and geometries most often used have been either S waves directly above earthquakes^{9,14} or ScS waves that reflect off the core-mantle boundary (CMB) from earthquakes less than ~4,000 km away¹⁵⁻¹⁸. But both require nearby seismogenic zones, and are thus ill-suited for studying continents which are typically devoid of earthquakes, except near their margins. There is another wave type, SKS, that travels through the mantle as an S wave and the outer core as a P wave, possessing a nearly vertical mantle path for even the most distant earthquakes (Fig. 1). SKS has not been exploited to the extent of ScS or S. To our knowledge the only quantitative analysis of this phase for splitting is for two nearby stations in Germany^{19,20}.

The present report represents the beginning of a systematic study of continental anisotropy using SKS phases recorded by various digital broadband networks. We have used a suite of observations to place constraints on the vertically integrated anisotropy from the CMB to the surface for nine continental stations in North America and Europe (Fig. 1). We find clear evidence of anisotropy at all stations, although both the magnitude, δt , defined as the delay time between the fast and slow phases, and the orientation of the fast polarization direction, ϕ , are found to change over distances of ~1,000 km. The strongest

Fig. 1 Map in Mercator projection of the nine stations used in this study. Size of circle gives magnitude of delay time δt between split shear wave arrivals and arrow gives azimuth of fast polarization direction. Partially filled circles denote those stations with only one measurement which should be regarded as preliminary. Shaded boundary marks southwestern extent of Canadian Shield. Inset: path of the phase SKS through the Earth. Stars denote location of earthquakes, triangle gives station location. Phase propagates as an S wave from earthquake to core-mantle boundary (CMB), through the outer core as a P wave, and again as an S wave from the CMB to the receiving station. Because the P-to-S conversion at the CMB removes all but radially polarized energy, SKS is only sensitive to splitting produced by anisotropy for the S leg underneath the station, and is thus ideal for high-resolution measurements of particular geologic terrains. This phase is well excited and isolated beyond 85° , and can thus be used in seismically inactive regions such as stable continental shields and platforms.



anisotropy, $\delta t = 1.75$ s, among the largest delay times ever obtained (compare for example refs 15–18), is found for a station in the Canadian Shield, which appears to reflect 'fossil' mantle deformation produced at the end of the Archean, as we will show.

Observations of transversely polarized SKS

Because SKS has been converted from a P wave to an S wave at the CMB, it will be polarized in the vertical plane of propagation for a spherically symmetric, isotropic Earth (Fig. 1). Thus, the transverse component of energy, SKS_T , will be identically zero. Both anisotropy and aspherical structure can produce a non-zero SKS_T through shear-wave splitting and perturbations to the ray path respectively. These two possibilities can be distinguished on the basis of particle-motion analysis. For aspherical structure, particle motion will remain predominantly rectilinear, whereas shear-wave splitting will produce elliptical particle motion. In the special case of SKS, for near-vertical propagation, and assuming transverse isotropy with a horizontal symmetry axis, the radial and transverse components $u_r(t)$ and $u_t(t)$, are simply related by

$$u_r(t) = s(t) \cos^2 \phi + s(t - \delta t) \sin^2 \phi \quad (1)$$

$$u_t(t) = -\frac{1}{2}[s(t) - s(t - \delta t)] \sin(2\phi) \quad (2)$$

where $s(t)$ is the radial waveform in the absence of anisotropy and ϕ is the angle between the fast and radial directions. For small δt compared to the dominant period of $s(t)$, the change in radial component will be subtle: a slight broadening and distortion. In contrast, the transverse component will be approximately proportional to the time derivative of the radial component (equation 2). In Fig. 2 is shown an example of a seismogram in which SKS_T is clearly observed and is about the same amplitude as SKS_R . Note that the particle motion is elliptical rather than rectilinear, and furthermore, that the displacement seismograms show the predicted time-derivative relationship. The other seismograms in our data set show the same kind of relationship, strongly suggesting that the transverse-component signals are predominantly caused by azimuthal anisotropy.

Method of measurement

We have searched for seismic events deeper than 100 km with well recorded SKS_R on the intermediate-period broad-band channel of the DWSSN, RSTN networks, and Grafenburg

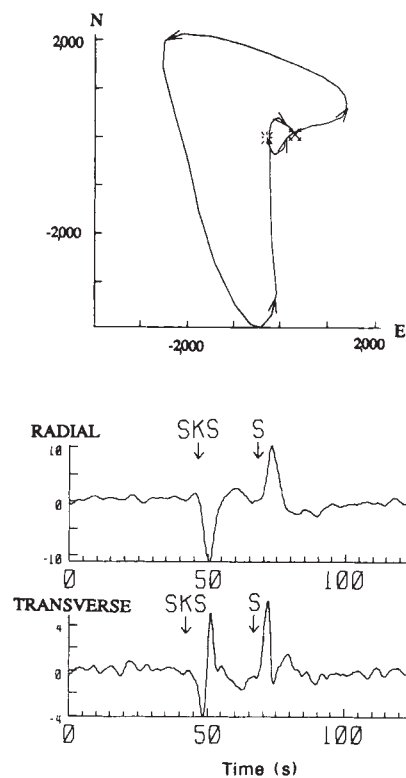


Fig. 2 Diagnostics of shear-wave splitting. The most general characteristic is elliptical particle motion. A seismogram at station RSON for an event in Izu-Bonin (10 September 1985; depth: 499 km, $m_b = 5.8$, distance = 89.1° , back azimuth = 314.3°) is shown. In the top panel is the particle motion in the horizontal plane (map view) for SKS. Crossed and uncrossed symbols mark the beginning and end of the phase, respectively, with the arrows giving the direction of time. Note that the initial polarization is slightly north of east and final polarization is closer to N-S, corresponding to the fast and slow directions respectively. For the phase SKS, because the polarization is in radial direction for an isotropic medium, an additional diagnostic is that the transverse component is non-zero and approximately the time derivative of the radial component (see equations 1 and 2 in text). Also shown are the (instrument-corrected) displacement records (bottom panel) which indicate the predicted derivative relationship.

Table 1 Shear-wave splitting measurements for the nine stations shown in Fig. 1

Station	N^* (SKS _R)	$N^\dagger$ (SKS _T)	$\phi^\ddagger$ (deg)	$\sigma_\phi^\S$ (deg)	$\delta t^\parallel$ (s)	$\sigma_{\delta t}^\S$ (s)	Latitude (deg)	Longitude (deg)	Location
RSON	11	8	75	1	1.75	0.10	50.86	-93.70	Red Lake, Ontario
RSNT	11	6	48	2	1.10	0.10	62.48	-114.59	Yellow Knife, N.W.T.
RSSD	12	4	54	8	0.60	0.10	44.12	-104.04	Black Hills, SD
COL	3	3	77	¶	1.55	¶	64.89	-147.79	College Outpost, Alaska
TOL	2	2	84	5	1.05	0.15	39.88	-4.05	Toledo, Spain
GRF	2	2	79	7	0.85	0.20	49.69	11.22	Bayern, FRG
RSCP	1	1	46	5	0.80	0.30	35.59	-85.57	Cumberland Plateau, TN
LON	1	1	-89	17	0.80	0.30	46.75	-121.81	Longmire, WA
KEV	1	1	01	23	1.15	0.80	69.76	27.01	Kevo, Finland
Total	44	28							

* Number of well observed SKS_R. † Number of well observed SKS_T. ‡ Azimuth of fast direction clockwise from north. § One standard deviation. ¶ Delay time. ¶ Inconsistent measurements, unweighted average taken (see text).

array (GRF). The best distance range for splitting measurements is 85°–110°, where SKS is both well isolated from S and ScS (>85°) and sufficiently energetic (<110°). In total, 44 SKS_R phases were obtained with 28 visually observable SKS_T. Although these 28 observations form the basis of our measurements, the absence of SKS_T for a well excited SKS_R provides the added constraint that either the anisotropy is small ($\delta t \sim 0$), or the back azimuth is along the fast or slow polarization direction (equation 2).

As is common in making splitting measurements, we search for the pair ϕ and δt that, when used to correct for the anisotropy, most successfully removes its effect¹⁴. Because SKS is initially radially polarized, this may be simply done by minimizing the energy $E_t(\phi, \delta t)$ on the reconstructed transverse component $\tilde{u}(t)$:

$$E_t(\phi, \delta t) = \frac{1}{T} \int_0^T \tilde{u}_t(t)^2 dt \sim \frac{1}{N} \sum_{n=1}^N (\tilde{u}_t^n)^2 \Delta t \quad (3)$$

for an N -point digital time series. E_t may be quickly evaluated for many candidate values of ϕ and δt (increments of 1 degree and 0.05 s respectively), as it can be expressed as a linear combination of the auto- and cross-correlations of the original E-W and N-S seismograms. The confidence region is found from the F -test and defined by values $(\phi, \delta t)$ that satisfy $E_t/E_t^{\min} < F_\alpha(\nu, \nu)$ where $E_t^{\min} = \min(E_t)$. $F_\alpha(\nu, \nu)$ is the critical value of the F -test for a confidence level of α , taken to be 0.05, and for ν degrees of freedom, representing the number of uncorrelated time points ($\leq N$). We emphasize that because of the high quality of the data, the use of the low-noise transverse component and the ability to assess uncertainties on individual measurements, it is possible to obtain a reliable estimate of $(\phi, \delta t)$ from one well recorded SKS_T seismogram.

In Fig. 3a and b, we illustrate the measurement procedure by presenting two original and reconstructed seismograms both for stations RSON and RSSD. The success of this two-parameter model in eliminating the transverse energy strongly suggests that we have correctly modelled the SKS_T phases. As a check of our estimates, we note that $(\phi, \delta t)$ for RSON is (75°, 1.75 s). As ϕ is nearly E-W, we should be able to take the original E-W, N-S seismograms and directly observe the anisotropy. Figure 3c shows these two components for two events. It is apparent that the E-W and N-S seismograms are very similar in shape and that the E-W component is early by between 1.5 and 2 s. In Fig. 3d are shown the eight individual measurements for RSON. Note that the 1σ confidence intervals appear to adequately reflect the scatter in the measurements.

The averages for the nine stations (weighted by the individual σ -estimates) are given in Table 1 and Fig. 1. For each station, we have between one and eight measurements of SKS_T, with values of δt ranging from 0.6 s for RSSD to 1.75 s for RSON. The lateral variations in $(\phi, \delta t)$ are striking; note, for example,

that these two extreme values are for the two stations in closest proximity (separated by about 1,000 km). The partially filled symbols in Fig. 1 correspond to those stations with only one SKS_T measurement (LON, RSCP, KEV), which should therefore be considered preliminary.

Localization of the anisotropy

Splitting in the phase SKS represents an integral average of the anisotropy along the second S-leg (Fig. 1) from the CMB to the surface. Although in principle the anisotropy could be anywhere along this path, there are two features of the data that suggest that it is near the surface. The first is the close proximity but dissimilarity of anisotropy for stations RSON and RSSD, whose delay times differ by more than 1 s (Table 1). As the paths for these stations are separated by less than 600 km at the CMB, strong, short-wavelength heterogeneity would be required there. On the other hand, the inferred anisotropy appears to be insensitive to back-azimuth in most cases. For example, in Fig. 3a the two records shown for RSON are for one event in the Western Pacific (event 1, back azimuth: 316.4°), and one in Hindu Kush (event 2, back azimuth: 12.4°). Although the emergent points at the CMB are separated by nearly 1,500 km, the two measurements give nearly the same value of δt (1.80 s versus 1.75 s) and ϕ (80° versus 82°). This similarity would thus require uniform CMB anisotropy on 1,500-km length scales. As these two properties are incompatible, it is concluded that the anisotropy is far removed from the core-mantle boundary and most plausibly localized near the surface.

Furthermore it appears that the upper mantle, rather than the crust, is the primary source of anisotropy. We first note that to confine the anisotropy entirely to the crust would require an unreasonably high value of 16% polarization anisotropy over the entire 40-km-thick crust under RSON. Thus, any crustal contribution would have to be coherent with the mantle (that is, with the same fast polarization direction); otherwise it could complicate or even diminish the mantle component. The shear-wave splitting measurements that exist for the upper crust⁹ and entire crust²¹ yield delay times in the range 0.1–0.3 s, roughly a factor of five smaller than what we observe, on average, suggesting that the crustal contribution is small.

A direct assessment of the total crustal component beneath RSON can be obtained by the use of the radially polarized teleseismic phase, PrnS, that converts from P to S at the Moho, as it is only sensitive to shear-wave splitting in the crust. At RSON in particular, this phase is clearly observable and has been used previously to model the crustal structure under this station²². We have searched for very impulsive events recorded at RSON, for which PrnS is clearly seen, and the back-azimuth is within 5° of where we see the strongest SKS_T arrivals (see for example Fig. 3a, top event at back azimuth of 316.4°). Two such events were found from the Sea of Okhotsk (20 April 1984, $m_b = 6.0$, depth = 581 km, back azimuth = 321°; 23 April 1984,

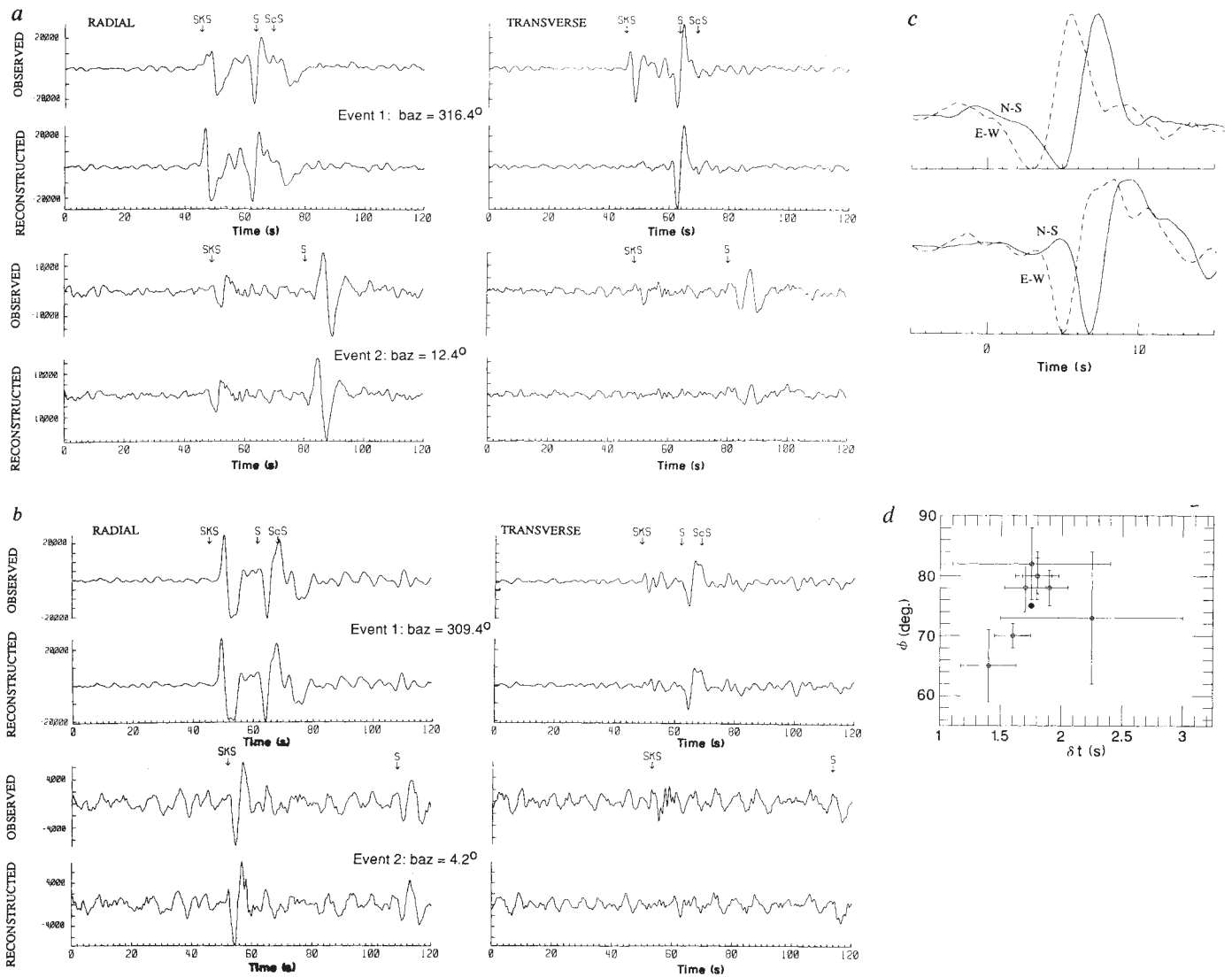


Fig. 3 Original and reconstructed radial and transverse seismograms from two earthquakes for two of the nine stations used in this study: *a*, RSON and *b*, RSSD. Event 1: 10 April 1985; Izu Bonin; depth = 398 km, $m_b = 5.8$. Event 2: 16 February 1984, depth = 208 km, $m_b = 6.1$. Back azimuth (baz) is defined as the azimuth (clockwise from north) of the earthquake with respect to the station. For SKS, this is also the polarization direction in an isotropic medium. Anisotropic parameters for individual measurements are: *a*, RSON top: $(\phi, \delta t) = (80 \pm 4^\circ, 1.80 \pm 0.20 \text{ s})$; bottom: $(82 \pm 6^\circ, 1.75 \pm 0.65 \text{ s})$; *b*, RSSD top: $(\phi, \delta t) = (52 \pm 20^\circ, 0.95 \pm 0.60 \text{ s})$; bottom: $(52 \pm 12^\circ, 0.60 \pm 0.15 \text{ s})$. Reconstructed seismograms have been rotated into the coordinate system defined by ϕ (the estimated fast direction) and its orthogonal, shifted by the negative of the delay time $-\delta t$ so as to remove the anisotropy. If successful, the transverse component should be identically zero, as in an isotropic Earth. Note that the estimated values of $(\phi, \delta t)$ remove nearly all of the transverse signs. *c*, Original E-W, N-S seismograms for two events at RSON. Top: 10 September 1985, Izu-Bonin, $m_b = 5.8$; bottom: event 1 in Fig. 3*a*. As ϕ is nearly E-W, anisotropy should be evident on original unrotated seismograms. Note that waveforms are very similar and the E-W component is obviously early by 1.5–2.0 s. *d*, Confidence intervals for the eight measurements at station RSON. Vertical axis is fast azimuth (clockwise from north, so that $\phi = 45^\circ$ is NE-SW and $\phi = 90^\circ$ is E-W). Horizontal axis is the delay time δt between fast and slow components. Each estimate of $(\phi, \delta t)$ is given by an open circle, with the 1σ confidence interval. The weighted average of all the measurements is shown as a filled circle. Note that the confidence intervals reliably indicate the scatter in the measurements.

$m_b = 6.0$, depth = 415 km, back azimuth = 321°). Whereas PmS is observed on the radial components of both events, the transverse components show no detectable signal during the same time period. This strongly suggests that the coherent crustal component is no more than ~ 0.5 s (corresponding to 4% anisotropy throughout the entire crust), and is probably closer to the 0.1–0.3 s range quoted above.

Assuming that the anisotropy is from a localized layer of mantle material with constant shear velocity β_0 , and that for a vertical ray path, the thickness of the layer L is approximately given by $L = \delta t \beta_0 / \epsilon$, where $\epsilon (\ll 1)$ is the fractional difference in velocity between the fast and slow polarization directions,

β_0 can easily be obtained from published velocity models²³, but ϵ must be constrained from more indirect considerations. First, we assume that the anisotropy is due to the preferred orientation of olivine and orthopyroxene, rather than, for example, the alignment of cracks, since cracks would be difficult to sustain below the shallow crust. The maximum single-crystal anisotropy expected from olivine, probably the most abundant and most anisotropic mineral in the upper mantle, is $\sim 10\%$ for shear-wave splitting, assuming (vertical) propagation along the c -axis (propagation along the b - and a -axes give about 7% and 2.5% respectively)^{24,25}. The presence of other minerals, particularly orthopyroxene, and the tendency for crystals to be only partially

aligned, cause the effective value to be somewhat lower. For a peridotite composition (60–80% olivine and the rest orthopyroxene), and using the velocity measurements of naturally occurring peridotites in ophiolites, the effective anisotropy is expected to be no greater than about 4% (ref. 26) or half the single-crystal value. Using $\beta_0 = 4.6 \text{ km s}^{-1}$, each second of δt thus corresponds to an anisotropic layer 115 km thick, assuming c -axis propagation, with proportionately greater thickness for other propagation directions. Applied to our data set, this gives a range of thicknesses from 70 km (RSSD) to 200 km (RSON). Allowing 0.2 s of delay time for crustal anisotropy would reduce these estimates by ~ 25 km.

Relation to tectonic processes

What is producing the anisotropy? In addressing this question, we focus on three stations in North America, RSON, RSNT and RSSD, for which we have the most measurements and which show significant variations in both magnitude and orientation over a relatively small region. It would be difficult to obtain the observed magnitude of anisotropy, unless the a -axis of olivine were in the horizontal plane and parallel to the fast polarization direction. As has been found in oceanic and continentally derived mantle peridotites and laboratory samples²⁷, there is a clear relationship between the orientation of the crystallographic axes and flow direction. For steady-state flow in the asthenosphere, the a -axis of olivine is expected to be aligned parallel to the flow plane and pointing in the flow direction. For continental orogenic deformation, the a -axis is expected to be aligned nearly parallel to the extension direction (more precisely, parallel to the foliation plane that approaches the extension direction for large strains), with the b -axis parallel to the direction of maximum shortening. The crystals of orthopyroxene are predominantly aligned with the olivine crystals (pyroxene a -, b - and c -axes are generally parallel olivine b -, c - and a -axes). Either deformation mechanism would thus put the a -axis of olivine in the horizontal plane. But in the case of RSON, RSSD and RSNT, the lateral variations in both the magnitude and orientation of the anisotropy over 1,000-km length scales would tend to support orogenic deformation, as asthenosphere flow would presumably produce coherent anisotropy throughout the North American plate.

The strains inferred from the anisotropy do not appear to be related to present-day crustal deformation. Crustal stress measurements show a coherent trend of E-W to NE-SW compression for Canada and the northern part of the United States^{28,29}, including the stations of interest, whereas the fast polarization directions imply more nearly N-S to NW-SE compression.

The magnitude and orientation of the anisotropy appear to be related to surface geology, however. The chief geological difference between RSON and RSSD is that one is in the predominantly Archean Canadian Shield (Superior Province), whereas the other is in the Wyoming Craton, which is primarily Proterozoic in age, or at least has been subject to extensive reworking in the Proterozoic. RSNT is also on the Canadian Shield, although it is in the Slave Province (Figs 1 and 4). Several studies have shown that the upper mantle under the Canadian Shield is seismically fast^{30–33}. The most detailed of these³³ reveals a strong gradient in upper-mantle velocities between RSON (high) and RSSD (low) and indicates that the Superior Province has the fastest velocities and deepest root in North America.

For the two shield stations, the orientation of the anisotropy is closely associated with late-Archean geological fabric. Namely, there is a striking correlation between the anisotropic fast polarization direction and the dominant grain present in large- and small-scale structures. RSON is situated within a 1,000 km by 800 km region of E-W-trending gneiss and greenstone terrains in the western Superior Province, as seen in Fig. 4. In the vicinity of RSON, located near the southern margin

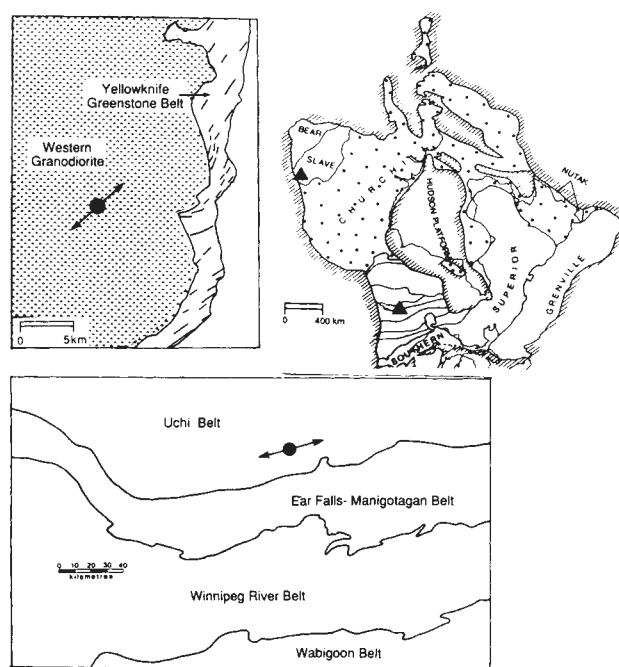


Fig. 4 Top right: map of the greenstone and gneiss belts of the Canadian Shield, and the locations of RSON, between the Uchi and English River Belts of the Western Superior Province, and RSNT adjacent to the Yellowknife belt in the Slave Province (after Franklin and Thorpe⁴⁴). Bottom: Detail of RSON location near boundary between Uchi belt and Ear Falls-Manitogagan belt (part of English River belt)⁴⁵. Arrow gives fast azimuth (75°) of the anisotropy. This is very close to the 83° average strike of the dominant metamorphic fabric (G. Stott, personal communication), as well as the boundary between belts. Fabric is the result of a 2.7 Gyr-old N-S compressional episode at the end of the Archean. Left: detail of the RSNT location just west of the Yellowknife greenstone belt in the Western Granodiorite⁴⁶. Dominant foliation direction in Yellowknife belt ranges from 30° to 50° . Foliation direction in the granodiorite is also NE-SW (ref. 34), suggesting that it was subjected to the same deformation. This direction closely corresponds to our estimate ϕ of 48° . The fabric is the result of a late-stage Archean NW-SE compressional episode.

of the Uchi Belt, the strike of the dominant structures is slightly north of east, with a mean strike azimuth of 83° (G. Stott, personal communication), which is only 7° different from our value of ϕ ($=75^\circ$). Furthermore, ϕ is virtually parallel to the boundaries between the belts (Fig. 4).

The same relationship appears to hold for RSNT in the Slave Province. Throughout the Slave Province is an important late-Archean metamorphic fabric with north-easterly trending cleavage, which, in the Yellowknife greenstone belt near RSNT, is manifested as foliations ranging between 30° and 50° in azimuth (Fig. 4). Our estimate of $\phi = 48^\circ$ shows remarkable agreement. Although the station is in the Western Granodiorite and not in the greenstone belt, the presence of northeast trending foliation in this unit as well, suggests that it has been subject to the same deformation³⁴. The roughly 25-km horizontal-averaging length of the SKS phases suggests that the Proterozoic deformation about 100 km to the southeast (for example, Great Slave Lake shear zone) would not contribute to the observed anisotropy. Thus for both of these stations, the anisotropy appears to have been caused by a compressional episode that marked the conclusion of an important period of Archean crust formation.

The geological history for RSSD (Black Hills, South Dakota) is more complicated, because it is located near the edge of the craton proximate to the trans-Hudson orogenic zone and was subject to several deformational episodes since the Archean³⁵. The dominant structural fabrics are less pervasively developed

and less coherent on the large scale, which may be responsible for the small value of δt .

Although we have fewer measurements for the other stations, it is worthwhile to briefly mention their relation to surface geology and tectonics. The station RSCP on the Cumberland Plateau (Fig. 1) shows a fast polarization direction parallel to the NE-SW trend of the Appalachian Mountains and appears to reflect the same compressional deformation that created this foldbelt. The two other North American stations, COL and LON, are situated near active convergent margins and are in the vicinity of subducted slabs. The E-W fast direction at LON may either reflect the present-day N-S crustal stress direction²⁹, or the fast direction in the subducted slab underneath the station. COL is probably the most complex geologically, with accretion-related Paleozoic and Cenozoic deformation³⁶, as well as the presence of the subducted Aleutian slab. It is also the station that gives clearly inconsistent measurements: two yield E-W fast, whereas a third gives NE-SW fast, the difference being much larger than the estimated uncertainties. In fact, this may reflect multiple sources of anisotropy.

In Europe, the fast polarization direction for station GRF ($\phi = 79^\circ$) is nearly parallel to the predominant Hercynian deformation, suggesting that it may be the primary cause of the anisotropic fabric. The published value of ϕ for GRF (ref. 20), using splitting in SKS, is nearly N-S, and thus inconsistent with our estimate. Their value was found to be in error, and the corrected value of nearly E-W fast is consistent with our results³⁷. The deformation in the vicinity of TOL was primarily caused by the Hercynian orogeny with relative tectonic stability since then. Our value of $\phi = 84^\circ$ roughly corresponds to E-W trend of the deformation³⁸. KEV, like RSON and RSNT, is in Archean terrain (Karelian province of the Baltic Shield) and thus may be used to compare seismic anisotropy with Archean deformational history. It is unfortunately the least well constrained of the nine stations; the preliminary N-S fast polarization direction is nevertheless parallel to the metamorphic fabric produced by a late Archean compressional episode³⁹.

Discussion

The close association between the fast anisotropic direction and the dominant geological fabric suggests that in many cases we are observing fossil strain associated with the last major orogenic episode. In the special case of the Canadian Shield, this would imply that the mantle fabric is 2.5–2.7 Gyr old and associated with late-stage Archean deformation. If this interpretation is correct, then several implications immediately follow. First, it will now be possible to use fossil seismic anisotropy, in the form of shear-wave splitting, to study the evolution and deep deformation

of continents back to the age of the oldest continents, some 3.8 Gyr ago. Second, the data set presently in hand suggests that at least part of the Canadian Shield (Superior Province) had already acquired, by the end of the Archean, a well developed 200–250-km-thick thermal boundary layer which has moved coherently with the surface for the past 2.5 Gyr. The survival of the mantle fabric over such a long period of time additionally argues for a cold subcontinental mantle (below $\sim 900^\circ\text{C}$) down to the deepest anisotropy, given the published data for diffusion in olivine⁴⁰. This would suggest a very low geothermal gradient of about 4°C km^{-1} beneath the Superior Province, a conclusion that is consistent with the observed high-seismic-velocity root discussed previously³³. The existence of such a well developed mantle root in the Archean argues against thermal-evolution models of the subcontinental mantle⁴¹, or models that advocate significant growth by underplating since the Archean; rather, the data would suggest that the root was rapidly formed by orogenic episodes at the end of the Archean.

Finally, the fact that these Archean, fabric-inducing, orogenic episodes were followed by more than 2.5 Gyr of tectonic stability suggests that they may be an integral part of craton-stabilization process. Significant mechanical stability is undoubtedly gained by the colder temperatures in the mantle roots, presumably introduced by advective thickening of the thermal boundary layer⁴². We speculate that an additional consequence of the deformation is the mechanical strengthening of the mantle by strain-hardening, a well known process in metals (as in the cold-rolling of steel) and silicates which strengthen materials by producing strains at temperatures that are low compared with the melting point. This occurs primarily by increasing dislocation density, but also leaves the fabric with a preferred crystal orientation, and is thought to be the dominant process for the deformation of the upper continental crust⁴³. This suggestion is potentially testable by comparison of the mechanical properties of deformed mantle xenoliths having raised dislocation densities with those of relatively undeformed samples.

We thank S. Shirey for help in placing the anisotropic measurements in a geological context, especially in the Canadian Shield, T. Kusky and W. Padgham for help with the geology of the Slave province near Yellowknife, and G. Stott and D. Troop for Red Lake geology. We also thank S. Crampin, N. Christensen, V. Babuška, J. Boyd, S. Karato, A. Nicolas and Eileen McLellan for discussion, and P. Shearer, N. Christensen, W. Padgham and N. J. Valette-Silver for comments on the manuscript. We finally thank Janice Dunlap for preparation of the manuscript. This work was supported by the Carnegie Institution of Washington, DTM and the NSF.

Received 25 February; accepted 13 June 1988.

- Hess, H. H. *Nature* **203**, 629–631 (1964).
- Raitt, R. W., Shor, G. G., Francis, T. J. G. & Morris, G. B. *J. geophys. Res.* **74**, 3095–3109 (1969).
- Shearer, P. M. & Orcutt, J. A. *Geophys. J. R. astr. Soc.* **87**, 967–1003 (1986).
- Forsyth, D. W. *Geophys. J. R. astr. Soc.* **43**, 103–162 (1975).
- Tanimoto, T. & Anderson, D. L. *J. geophys. Res.* **90**, 1842–1858 (1985).
- Bamford, D. *Geophys. J. R. astr. Soc.* **49**, 29–48 (1977).
- Bamford, D., Jentsch, M. & Prodehl, C. *Geophys. J. R. astr. Soc.* **57**, 397–429 (1979).
- Vetter, U. & Minster, J.-B. *Bull. seism. Soc. Am.* **71**, 1511–1530 (1981).
- Crampin, S. & Booth, D. C. *Geophys. J. R. astr. Soc.* **83**, 75–92 (1985).
- Booth, D. C., Crampin, S., Evans, R. & Roberts, G. *Geophys. J. R. astr. Soc.* **83**, 61–73 (1985).
- Crampin, S., Peacock, S. & Chen, T.-C. *Trans. Am. geophys. Un.* **67**, 1117 (1986).
- Dziewonski, A. M. & Anderson, D. L. *J. geophys. Res.* **88**, 3295–3314 (1983).
- Keith, C. M. & Crampin, S. *Geophys. J. R. astr. Soc.* **49**, 225–243 (1977).
- Bowman, J. R. & Ando, M. *Geophys. J. R. astr. Soc.* **88**, 25–41 (1987).
- Ando, M. & Ishikawa, Y. *J. phys. Earth* **30**, 191–199 (1982).
- Ando, M., Ishikawa, Y. & Yamazaki, F. *J. geophys. Res.* **88**, 5850–5864 (1983).
- Ando, M. *J. phys. Earth* **32**, 179–196 (1984).
- Fukao, Y. *Nature* **309**, 695–698 (1984).
- Vinnik, L. P., Kosarev, G. L. & Makeyeva, L. I. *Dokl. Akad. Nauk USSR* **278**, 1335–1339 (1984).
- Kind, R., Kosarev, G. L., Makeyeva, L. I. & Vinnik, L. P. *Nature* **318**, 358–361 (1985).
- Kaneshima, S., Ando, M. & Kimura, S. *Nature* submitted.
- Owens, T. J., Taylor, S. R. & Zandt, G. *Bull. seism. Soc. Am.* **77**, 631–662 (1987).
- Grand, S. & Helmberger, D. *Geophys. J. R. astr. Soc.* **76**, 399–438 (1984).
- Verma, R. K. *J. geophys. Res.* **65**, 757–766 (1960).
- Kumazawa, M. & Anderson, O. L. *J. geophys. Res.* **74**, 5961–5972 (1969).
- Christensen, N. I. & Lundquist, S. M. *Geol. Soc. Am. Bull.* **93**, 279–288 (1982).
- Nicolas, A. & Christensen, N. I. in *Composition, Structure and Dynamics of the Lithosphere-Asthenosphere System* Vol. 16 (eds Fuchs, K. & Froidevaux, C.) (Am. Geophys. Un., Washington D.C., 1987).
- Adams, J. *Earth Physics Open File* 85–31 (1986).
- Zoback, M. L. & Zoback, M. D. in *Geophysical Framework of the Continental United States* (eds Pakiser, L. & Mooney, W.) (Geol. Soc. Am., Boulder, Colorado) (in the press).
- Brune, J. & Dorman, J. *Bull. seism. Soc. Am.* **53**, 167–210 (1963).
- Woodhouse, J. H. & Dziewonski, A. M. *J. geophys. Res.* **89**, 5953–5986 (1984).
- Wickens, A. J. & Buchbinder, G. G. R. *Bull. seism. Soc. Am.* **70**, 809–822 (1980).
- Grand, S. P. *J. geophys. Res.* **92**, 14065–14090 (1987).
- Henderson, J. B. *Geol. Surv. Can.* **414**, 63–68 (1985).
- Dewitt, E., Redden, J. A., Wilson, A. B. & Buscher, D. *US Geol. Sur. Bull.* **1580**, 135 (1986).
- Jones, D. L., Silberling, N. J., Gilbert, W. & Consey, P. *J. geophys. Res.* **87**, 3709–3717 (1982).
- Vinnik, L. P., Makeyeva, L. I., Kosarev, G. L. & Kind, R. *Proc. IUGG Gen. Assem. 19th IASPEI Symp.*, 1:330 (1987).
- Banda, E., Surinac, E., Aparicio, A., Sierra, J. & Ruiz de la Parte, E. *Geophys. J. R. astr. Soc.* **67**, 779–789 (1981).
- Gaál, G., Mikkola, A. & Söderholm, B. *Precamb. Res.* **6**, 199–215 (1978).
- Goetze, C. & Kohlstedt, D. L. *J. geophys. Res.* **78**, 5961–5971 (1973).
- Sclater, J. G., Jaupart, C. & Galson, D. *Rev. Geophys. Space Phys.* **18**, 269–311 (1980).
- Jordan, T. H. *Phil. Trans. R. Soc. A301*, 359–373 (1981).
- Nicolas, A. & Poirier, J. P. *Crystalline Plasticity and Solid State Flow in Metamorphic Rocks* (Wiley, New York, 1976).
- Franklin, J. M. & Thorpe, R. I. in *H. S. Robinson Memorial Volume* (eds Hutchinson, R. W., Spence, C. D. & Franklin, J. M.) (Geol. Ass. Can. special paper 25, 1982).
- Beakhouse, G. P. in *Evolution of Archean Supracrustal Sequences* (eds Ayres, L. D., Thurston, P. C., Card, K. D. & Weber, W.) 169–178 (Geol. Ass. Can. special paper 28, 1985).
- Fyson, W. K. *Can. J. Earth Sci.* **19**, 878–893 (1982).

Repetitive spikes in cytoplasmic calcium evoked by histamine in human endothelial cells

Ron Jacob, Janet E. Merritt, Trevor J. Hallam & Timothy J. Rink

Smith Kline & French Research Limited, The Frythe, Welwyn, Hertfordshire AL6 9AR, UK

Measurement of cytoplasmic free calcium, $[Ca^{2+}]_i$, in single human endothelial cells has shown that low doses of the inflammatory mediator histamine evoke asynchronous repetitive spikes in $[Ca^{2+}]_i$ whereas high doses cause a maintained elevated $[Ca^{2+}]_i$. We discuss possible regulatory mechanisms, and the potential physiological and pathological implications of such a frequency-modulated $[Ca^{2+}]_i$ signalling system.

HISTAMINE was the first identified mediator of inflammation, and histamine (H_1) antagonists are effective in moderating allergic and acute inflammatory responses¹. One of the primary sites of action is on post-capillary venule endothelial cells where it causes rapid changes in vascular permeability partly, at least, by cell retraction². Histamine and several other inflammatory mediators have been shown to promote hydrolysis of phosphatidylinositol (4,5)-bisphosphate and to increase cytosolic Ca^{2+} ($[Ca^{2+}]_i$) in populations of endothelial cells³⁻⁵. These effects of histamine are blocked by mepyramine (an H_1 antagonist) but not cimetidine (an H_2 antagonist), showing that histamine is acting via H_1 receptors. Here we report responses to histamine of single fura-2-loaded endothelial cells from human umbilical veins. At low concentrations (0.1–3.0 μ M) histamine evoked a characteristic spiking $[Ca^{2+}]_i$ response with an initial increase to around 1 μ M and then repetitive smaller spikes whose frequency (0.3–1.5 min^{-1}) increased with increasing concentration of histamine. This pattern is reminiscent of that reported for aequorin-injected hepatocytes stimulated with vasopressin, angiotensin or phenylephrine^{6,7}. Because fura-2 can measure $[Ca^{2+}]_i$ at and below basal levels, we could see that $[Ca^{2+}]_i$ reverted to the basal level during each interspike interval and rose progressively in a 'pacemaker'-like way before the sharp upstroke. Spiking persisted in the continued presence of agonist for up to 40 min and was virtually unaffected by K^+ depolarization, eliminating an involvement of membrane potential in the mechanism of spike generation. At the highest histamine concentrations (10–100 μ M), the initial spike was followed by a plateau phase which sometimes fluctuated. In the absence of external Ca^{2+} , the $[Ca^{2+}]_i$ spikes progressively decayed in size and the interspike interval lengthened: readdition of Ca^{2+} restored the spike amplitude and frequency. Because endothelial cells lining blood vessels are normally in close contact, we used fluorescence imaging to measure $[Ca^{2+}]_i$ simultaneously in a field of several endothelial cells in a confluent monolayer and found that adjacent cells respond asynchronously to low doses of histamine. These data from intact human cells provide a striking example of frequency-modulated $[Ca^{2+}]_i$ signalling in a non-excitable cell, and give some clues about the mechanisms underlying the spike generation and the interspike interval.

Results

Response to histamine. Figure 1 shows the effects of histamine on single cells in the presence of 1-mM Ca^{2+} . A low dose of histamine (0.3 μ M) caused an increase in $[Ca^{2+}]_i$ from the basal level of 100 nM to a peak of $\sim 1 \mu$ M, which then rapidly declined back to the basal level. Thereafter, $[Ca^{2+}]_i$ continued to spike regularly at a frequency of 0.7 min^{-1} . This behaviour was seen in each of 40 histamine-responsive cells exposed to low doses of histamine (a few cells, less than 10%, were completely unresponsive to high or low doses of histamine). Figure 1b shows the responses of two different cells to increasing concentrations

of histamine. Low concentrations of histamine evoked repetitive $[Ca^{2+}]_i$ spikes whose frequency increased with increasing histamine concentration until, at high concentrations, the spikes apparently fused, resulting in a maintained elevated $[Ca^{2+}]_i$. Dose-dependent spiking was seen in each of 10 cells examined in this way although the sensitivity to histamine and the shape of the transients varied somewhat from cell to cell. The time lag of the initial response decreased with increasing agonist concentration, possibly reflecting the decreasing time required to generate a sufficient concentration of inositol (1,4,5)-trisphosphate ($\text{Ins}(1,4,5)\text{P}_3$) to discharge the internal Ca^{2+} pool. When the concentration of histamine was changed without any period of recovery between doses (Fig. 1c) the spiking frequency changed immediately on changing the concentration of histamine, suggesting that the response was dependent on the extent of receptor-occupation and was not related to the size and nature of the previous rises in $[Ca^{2+}]_i$. A similar dose-dependent modulation of spiking frequency has been reported for single aequorin-loaded hepatocytes stimulated with vasopressin and other ligands^{6,7}.

In contrast to aequorin, fura-2 can resolve changes in $[Ca^{2+}]_i$ close to the basal level. An expanded trace of the response to 0.3 μ M histamine (Fig. 1d) shows that a pattern reminiscent of a 'pacemaker' potential was often seen: between spikes $[Ca^{2+}]_i$ returned to the basal level and then progressively increased before the sudden upstroke of the next spike. This suggests the possibility that $[Ca^{2+}]_i$ might itself be involved in a regenerative process that leads to spiking.

Effect of high K^+ . The concept of transient repetitive signalling and frequency modulation is rather new for non-excitable cells^{6,7}; however, it is common in neuronal and muscle cell signalling, where it is generated by voltage-gated channels and action potentials^{8,9}. It was therefore important to see whether the repetitive spiking that we observed was related to membrane potential changes. Direct electrophysiological measurements are technically difficult in human endothelial cells so we took the simple approach of treating the cells with isotonic K^+ medium to hold the membrane potential close to 0 mV. (Previous studies with populations of endothelial cells have shown that K^+ depolarization does not itself alter resting or histamine-stimulated $[Ca^{2+}]_i$, arguing against the presence of voltage-gated Ca^{2+} channels³.) Exposure to high K^+ had little effect on a repetitively spiking cell (Fig. 2), strongly suggesting that the membrane potential plays no important role in generating the spikes.

Effect of removing external Ca^{2+} . Next, we investigated the influence of external Ca^{2+} to determine whether spike generation requires Ca^{2+} influx. Histamine was still able to evoke dose-dependent spikes of $[Ca^{2+}]_i$ in the absence of external Ca^{2+} , although both the frequency and amplitude of the spikes declined during stimulation (Fig. 3a), this decline being faster at the higher dose of histamine. These results show that the $[Ca^{2+}]_i$ responses are primarily due to discharge of an internal pool, presumably triggered by $\text{Ins}(1,4,5)\text{P}_3$. They also suggest

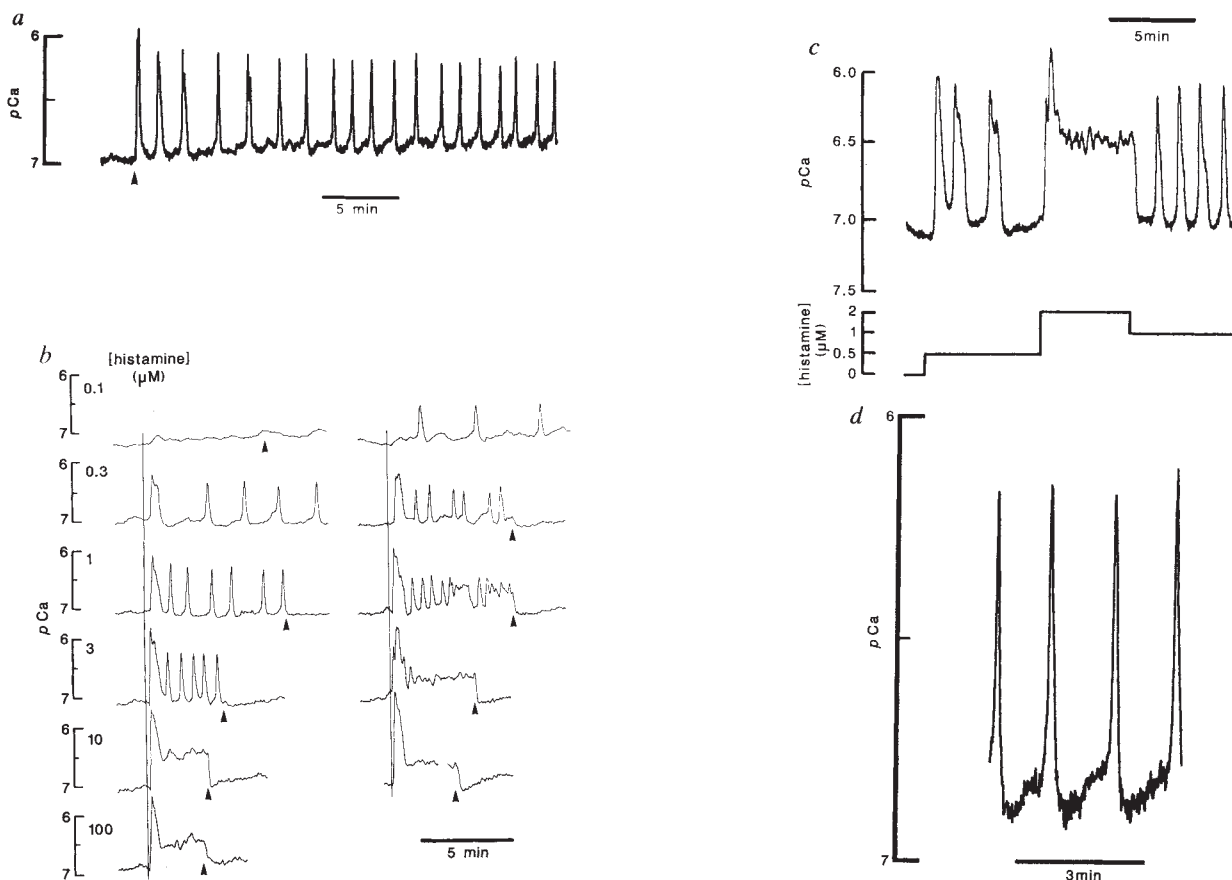


Fig. 1 *a*, The effect of histamine ($0.3 \mu\text{M}$) on $[\text{Ca}^{2+}]_i$ in a single human endothelial cell in the presence of 1 mM Ca^{2+} . The arrow indicates addition of histamine. *b*, Dose-dependent effects of histamine on $[\text{Ca}^{2+}]_i$ in the presence of 1 mM Ca^{2+} . Each column shows data from one cell exposed to a range of histamine concentrations. The interval between the removal of one dose of histamine and the application of the next dose was $\sim 3 \text{ min}$. The thin vertical line running through the start of each trace shows the time when the perfusion valve was switched to add histamine. The arrow, when present, shows the time when the perfusion valve was switched to remove histamine. There is a lag between switching the valve and the change of solution at the cell. This lag is constant for a particular cell. Blank sections of trace are where artefacts (due to bubbles) have been removed. *c*, Effect on $[\text{Ca}^{2+}]_i$ of changing histamine concentration without an intervening histamine-free phase. *d*, An expanded section of the trace from *a* showing the 'pacemaker'-like changes in $[\text{Ca}^{2+}]_i$ between transients.

Methods. Endothelial cells were obtained from human umbilical cords, as previously described³⁴, and were cultured (in 25-cm^2 flasks) in medium 199 (Flow Laboratories) containing 20% fetal calf serum (FCS), $50 \mu\text{U ml}^{-1}$ penicillin and $50 \mu\text{g ml}^{-1}$ streptomycin for ~ 1 week. Cells were then detached by exposure to trypsin (0.05%), reseeded (in the same medium) on to number-1 glass coverslips, and kept in culture for 3–4 days before use. To measure $[\text{Ca}^{2+}]_i$, cells were incubated in Hepes (20 mM)-buffered Dulbecco's-modified Eagles Medium (Flow Laboratories) with 20% FCS and $1\text{-}\mu\text{M}$ fura-2 AM (Molecular Probes) for 30 min at room temperature, and then placed in the superfusion medium supplemented with 1% bovine serum albumin and kept at room temperature for at least 30 min before use. A coverslip was placed on the thermostatted stage on an inverted microscope (Zeiss Axiomat) equipped with a $\times 40$ (Nikon) fluorite objective, and the cells were superfused at 0.3 ml min^{-1} with medium containing 145-mM NaCl , 5-mM KCl , 1-mM CaCl_2 , 1-mM MgCl_2 , 10-mM glucose , 0.1% bovine serum albumin and 10-mM Hepes ($\text{pH } 7.4$) at 37°C . Superfusion media were switched (for example, for addition of agonist) using miniature solenoid valves with a $100\text{-}\mu\text{l}$ volume. With the end of the perfusion line $1\text{--}6 \text{ mm}$ from the cell of interest, the solution-change at the cell surface was 90% complete within 7 s . The cells were illuminated via the epifluorescence port of the microscope and a 425-nm dichroic mirror, with alternating 340-nm and 380-nm excitation light from a Spex dual excitation wavelength fluorimeter. Fluorescence was collected via a 435-nm -long pass filter, detected with a photon counting photomultiplier driven from the Spex controller and counted in 1-s periods. An autofluorescence correction was made at the end of each run by exposing the cells to 2-mM MnCl_2 and $0.5\text{-}\mu\text{M ionomycin}$ ³⁴ to quench the fura-2 fluorescence. $[\text{Ca}^{2+}]_i$ was calculated from the ratio of the fluorescence at the two excitation wavelengths³⁵.

that the mechanism that triggers the spiking is independent of an extracellular source of Ca^{2+} , although extracellular Ca^{2+} is required to maintain the spiking, perhaps by replenishing internal Ca^{2+} stores or by contributing to the 'pacemaker' (thus influencing the frequency of spiking). Withdrawing external Ca^{2+} during an established response caused an immediate decline in both frequency and amplitude of the spikes (Fig. 3*b*); readdition of Ca^{2+} immediately restored the initial spiking pattern, showing that Ca^{2+} -free medium does not cause cell damage or desensitization to histamine.

Figure 4 shows experiments designed to investigate whether the declines in spiking frequency and amplitude in Ca^{2+} -free medium were associated with a depletion of the internal Ca^{2+} store. After incubating cells for $\sim 7 \text{ min}$ in Ca^{2+} -free medium, in the absence of histamine (as a control), the state of the internal

Ca^{2+} store was tested by application of a supramaximal dose of histamine ($100 \mu\text{M}$) still in Ca^{2+} -free medium (Fig. 4*a*). The response was similar to that normally produced by $100\text{-}\mu\text{M}$ histamine in Ca^{2+} -free medium without a prolonged preincubation (for example, Fig. 4*d*) and was also similar to the response that could subsequently be evoked after the internal Ca^{2+} stores had been refilled by exposing the cells to medium containing 1-mM Ca^{2+} for 3 min . When cells were incubated with a low dose of histamine ($0.15 \mu\text{M}$) in Ca^{2+} -free medium, only one spike in $[\text{Ca}^{2+}]_i$ was seen during the 7-min period (Fig. 4*b*). When the state of the internal Ca^{2+} store was subsequently tested by application of $100\text{-}\mu\text{M}$ histamine, still in Ca^{2+} -free medium, followed by a refilling period and a further application of $100\text{-}\mu\text{M}$ histamine, the two responses to $100\text{-}\mu\text{M}$ histamine were very similar, showing that the initial exposure to $0.15 \mu\text{M}$ his-

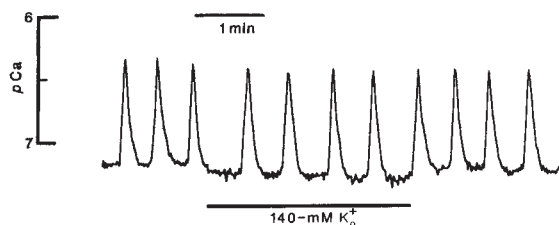


Fig. 2 Lack of effect of high K^+ solution on $[Ca^{2+}]_i$ spiking stimulated by $0.3\text{-}\mu\text{M}$ histamine. The high K^+ solution was the same as the normal perfusate except that $[K^+]$ was 140 mM and $[Na^+]$ was 10 mM . Similar results were obtained with four different cells.

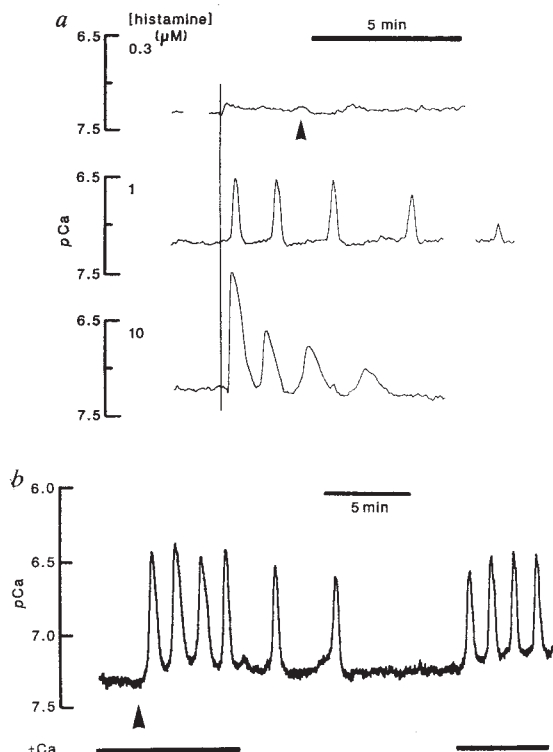


Fig. 3 *a*, The effect of histamine on $[Ca^{2+}]_i$ in one endothelial cell in the absence of extracellular Ca^{2+} . The perfusate was the same as for Fig. 1*a* except that $CaCl_2$ was omitted and 0.1-mM EGTA was added. In between each dose of histamine, internal stores of Ca^{2+} were refilled by a 3-min exposure to 1-mM Ca^{2+} in the absence of agonist. (Cells were generally less stable in Ca^{2+} -free solution so that only two or three doses of histamine could be tested on one cell.) *b*, Effect of removing external Ca^{2+} while a single endothelial cell was spiking in response to $0.5\text{-}\mu\text{M}$ histamine. A cell was perfused with medium containing either 1-mM Ca^{2+} or 1-mM EGTA, as shown. The arrow indicates the time of histamine addition; histamine was then present throughout. Similar results were obtained with six cells.

tamine had not caused any significant depletion of the internal Ca^{2+} store. But incubating cells with an intermediate concentration of histamine ($5\text{ }\mu\text{M}$) for 7 min evoked decreasing spikes in $[Ca^{2+}]_i$ (Fig. 4*c*). When the state of the internal Ca^{2+} store was then tested by application of $100\text{-}\mu\text{M}$ histamine, a moderate spike in $[Ca^{2+}]_i$ was seen, which was much smaller than that produced by $100\text{-}\mu\text{M}$ histamine after a refilling period. Thus, the few spikes in $[Ca^{2+}]_i$ stimulated by $5\text{-}\mu\text{M}$ histamine in Ca^{2+} -free solution partially, but not totally, depleted the histamine-sensitive internal Ca^{2+} store. A supramaximal dose of

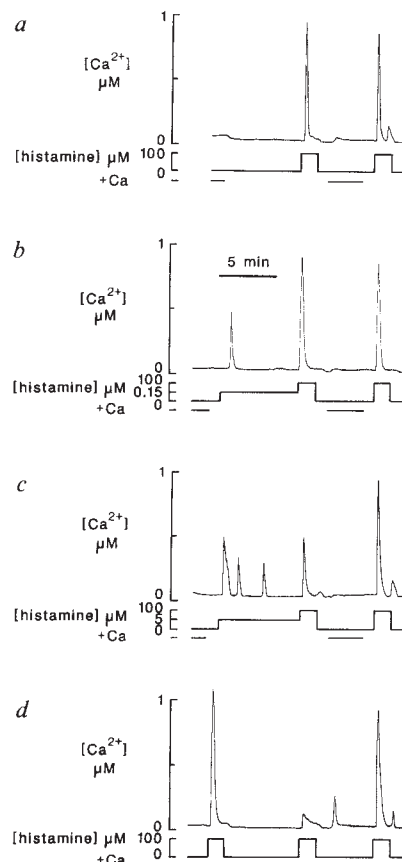


Fig. 4 The state of the internal Ca^{2+} store after exposure to various doses of histamine. Cells were incubated for 7 min in Ca^{2+} -free solution containing the following concentrations of histamine: 0 (*a*), $0.15\text{ }\mu\text{M}$ (*b*), $5\text{ }\mu\text{M}$ (*c*), or $100\text{ }\mu\text{M}$ for 1.5 min (*d*). The state of the internal Ca^{2+} store was then tested by application of $100\text{-}\mu\text{M}$ histamine still in Ca^{2+} -free medium (test response 1), compared to that evoked by $100\text{-}\mu\text{M}$ histamine applied after the internal Ca^{2+} stores had been refilled by exposure to 1-mM external Ca^{2+} for 3 min (test response 2). The relation between test responses 1 and 2 was significantly different in *a* and *c* in three experiments ($P = 0.05$, Mann and Whitney U test).

histamine ($100\text{ }\mu\text{M}$) completely discharged the histamine-releasable store of Ca^{2+} (Fig. 4*d*). In Ca^{2+} -free solution $100\text{-}\mu\text{M}$ histamine evoked one large spike which virtually depleted the internal pool of Ca^{2+} (Fig. 4*d*), because a subsequent application of $100\text{-}\mu\text{M}$ histamine in the continued absence of external Ca^{2+} evoked only a minimal response. Exposure to 1-mM external Ca^{2+} for 3 min was sufficient to allow refilling of the internal store because $100\text{-}\mu\text{M}$ histamine, in Ca^{2+} -free medium, then evoked a full response. During the refilling period, the elevation in $[Ca^{2+}]_i$ was virtually absent in some cells (for example, 15 nM in Fig. 4*c*) although in others it was pronounced (for example, Fig. 4*d*). This suggests that in some cells Ca^{2+} refills directly into the dischargeable store^{10,11} although this may not be the only mechanism.

Imaging of $[Ca^{2+}]_i$ in confluent cells. Under normal conditions endothelial cells lining blood vessels are in close contact, raising the possibility of intercellular communication. It is therefore of interest to see whether or not adjacent cells respond synchronously to histamine. Figure 5 shows the effect of histamine ($1\text{ }\mu\text{M}$, a concentration which stimulates $[Ca^{2+}]_i$ spiking in single cells; see Fig. 1) on $[Ca^{2+}]_i$ simultaneously measured in a field of seven endothelial cells in the presence of extracellular Ca^{2+} (1 mM). When the superfusate (initially histamine-free) was switched to one containing histamine ($1\text{ }\mu\text{M}$), $[Ca^{2+}]_i$ increased

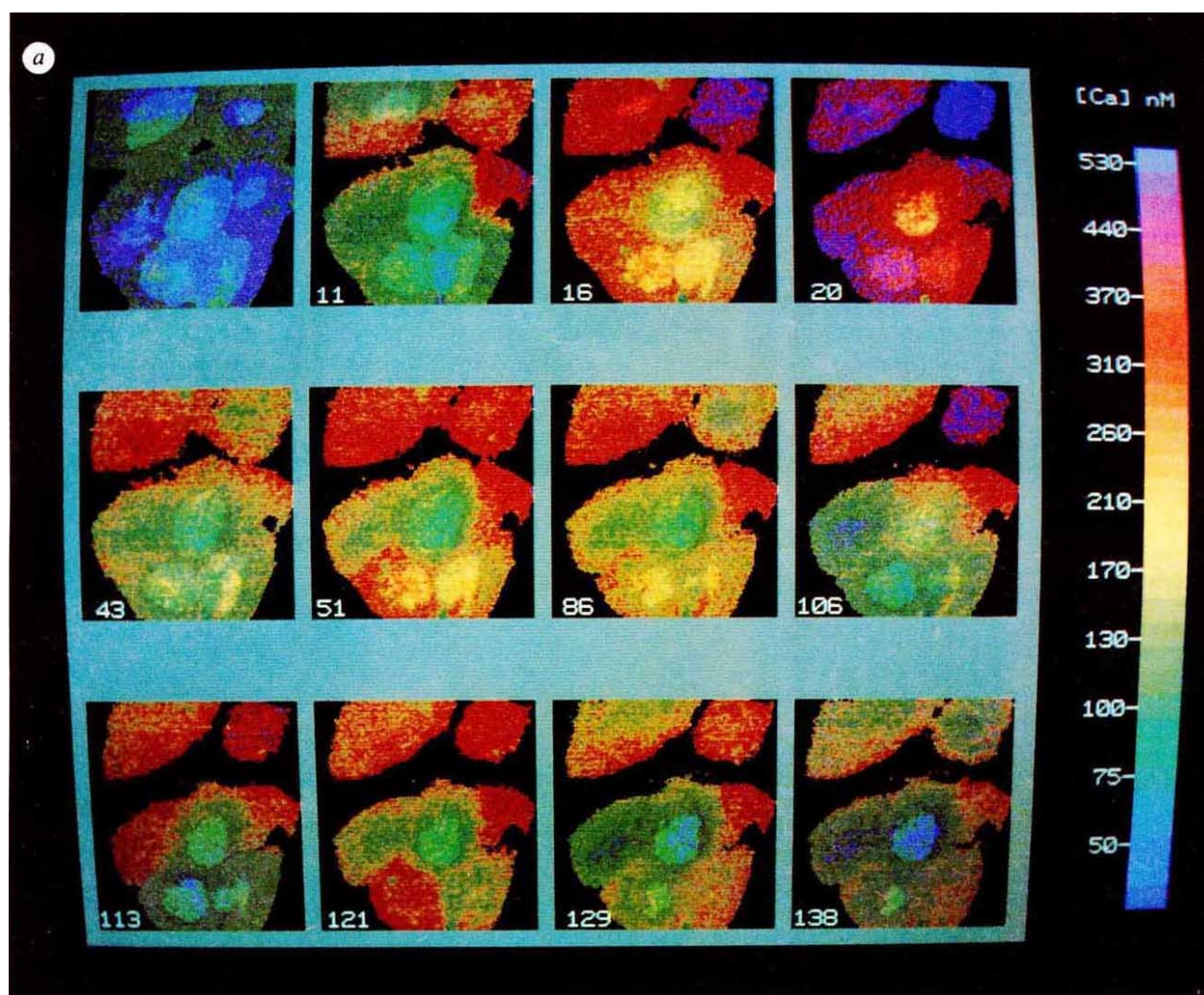


Fig. 5 *a*, $[Ca^{2+}]_i$ images of a confluent layer of endothelial cells stimulated with histamine ($1 \mu M$) in the presence of Ca^{2+} ($1 mM$). The scale on the right-hand side shows the colours corresponding to each $[Ca^{2+}]_i$. The numbers in the lower-left-hand corner of each frame indicate the time(s) after switching the superfusion to histamine. The superfusion was switched back to histamine-free solution at 132 s. *b*, Outline drawing delineating the positions of individual cells.

Methods. The $[Ca^{2+}]_i$ imaging was carried out according to the method of Tsien and Poenie³⁶. Cells were superfused and illuminated as above, but the fluorescence images were collected by a silicon intensified target camera (Cohu) and an image processor (Gould FD 5000) driven by a computer (DEC $\mu VAXII$) using software written by R. Y. Tsien. The images were digitized at 512×512 pixels. Averages of 16 frames were taken at each of the two excitation wavelengths (350 and 380 nm) and ratioed after background subtraction. This ratio image was then corrected for inhomogeneous shading, converted to $[Ca^{2+}]_i$, and displayed in false colour. When ratioing, pixels with an intensity below a threshold value are rejected (that is, the corresponding point on the ratio image is displayed as black). Because the 380-nm image dims when $[Ca^{2+}]_i$ rises, some pixels at the peripheries of the cells fall below the threshold value: this gives the impression that the cells are shrinking on stimulation, a condition which, although it may be true, cannot be detected using this technique.

in all of the cells and peaked at around 20 s after switching superfusates. Some cells were slower to respond than others; the differing latency periods may reflect different sensitivities of each cell to histamine. $[Ca^{2+}]_i$ then declined to 100–150 nM in all but one cell (number 1) after 43 s. Asynchronous transient elevations in $[Ca^{2+}]_i$ could be observed in various cells at 51 s (numbers 2 and 6), 106 s (number 2), 113 s (number 3), and 121 s (number 6). When the superfusate was switched back (at 132 s) to one still containing 1-mM Ca^{2+} but no histamine, the $[Ca^{2+}]_i$ spiking stopped and $[Ca^{2+}]_i$ fell back towards the basal pre-stimulated level (138 s). In one cell (number 1) $[Ca^{2+}]_i$ appeared to remain elevated for as long as histamine was present. This particular cell may have been more sensitive to histamine,

as this maintained elevation in $[Ca^{2+}]_i$ is more reminiscent of that seen in cells exposed to higher concentrations of histamine. These results show that contiguous cells responded asynchronously. Independent $[Ca^{2+}]_i$ spikes in contiguous cells have also been seen in BC3H-1 cells stimulated with phenylephrine or histamine¹².

Discussion

What causes the spiking behaviour? These findings, like those of Cobbold and his colleagues^{6,7}, prompt us to ask how a constant concentration of agonist produces repetitive spikes of $[Ca^{2+}]_i$. Our results provide some constraints for possible models. (1) Histamine can generate repetitive responses in the absence of

external Ca^{2+} , showing that the basic spiking mechanism is intracellular. (2) The response to histamine is unaffected in medium containing high $[\text{K}^+]$ which makes it very unlikely that $[\text{Ca}^{2+}]_i$ spiking is driven by fluctuations in membrane potential. Also, the lack of effect of the low external $[\text{Na}^+]$ in this medium makes it unlikely that $\text{Na}^+:\text{H}^+$ or $\text{Na}^+:\text{Ca}^{2+}$ exchange are important in this response. (3) There is a dose-dependent latency in the effect of agonist before a sharp rise in $[\text{Ca}^{2+}]_i$ is seen (also observed in hepatocytes^{6,7}), suggesting a threshold for the action of $\text{Ins}(1,4,5)\text{P}_3$ on net Ca^{2+} release from the internal store. A threshold dose for an effect of intracellularly microinjected $\text{Ins}(1,4,5)\text{P}_3$ on Ca^{2+} release (detected indirectly via a change in membrane potential) has been observed in Hamster eggs¹³. (4) The declining size of successive spikes in Ca^{2+} -free solution suggests that the state of the internal store is a determining factor in the spiking mechanism: this decline in response seems to be associated with a depletion of the internal Ca^{2+} store. (5) There appears to be a slow 'pacemaker' rise in $[\text{Ca}^{2+}]_i$ that may trigger the spike. (6) The spike height and duration are relatively independent of frequency of spiking, suggesting that the 'pacemaker' mechanism is independent of the mechanism for generating the spikes. (7) In Ca^{2+} -free solution, the frequency of spiking decreases, suggesting that there may be a Ca^{2+} -influx component of the pacemaker rise in $[\text{Ca}^{2+}]_i$. (This occurs at a time when the internal Ca^{2+} store is not significantly depleted.)

Spiking of $[\text{Ca}^{2+}]_i$ might be generated by spiking levels of $\text{Ins}(1,4,5)\text{P}_3$ (as proposed by Woods *et al.*⁷), perhaps generated by feedback loops involving Ca^{2+} . Alternatively, spiking might be generated by Ca^{2+} feedback loops operating in the presence of a constant level of $\text{Ins}(1,4,5)\text{P}_3$.

There are several mechanisms which could be involved in generating fluctuating levels of $\text{Ins}(1,4,5)\text{P}_3$ and hence $[\text{Ca}^{2+}]_i$ spikes. (1) Feedback inhibition by Ca^{2+} on the production of $\text{Ins}(1,4,5)\text{P}_3$. But there is no evidence that Ca^{2+} inhibits G-protein-activated phospholipase $\text{C}^{14,15}$. (2) Ca^{2+} could promote $\text{Ins}(1,4,5)\text{P}_3$ formation by synergizing with receptor/G-protein/phospholipase C complexes or by a direct effect on phospholipase C. This is a possibility because increasing Ca^{2+} from 0.1 to 1 μM is reported to potentiate G-protein stimulation of phospholipase $\text{C}^{14,15}$. (3) Inhibition by protein kinase C of agonist-stimulated phospholipase C: protein kinase C has been shown to inhibit G-protein-activated phospholipase C^{15} . (4) Ca^{2+} activation of $\text{Ins}(1,4,5)\text{P}_3$ kinase^{16,17}. This could have a negative effect by promoting the removal of $\text{Ins}(1,4,5)\text{P}_3$ ($\text{Ins}(1,4,5)\text{P}_3$ will also be removed by the lower-affinity phosphatase activity which is not regulated by Ca^{2+} in most tissues¹⁸) or a positive effect via the product of the phosphorylation, inositol (1,3,4,5)-tetrakisphosphate, which has been proposed to synergize with $\text{Ins}(1,4,5)\text{P}_3$ to facilitate Ca^{2+} movements^{19,20}.

Intracellular application of $\text{Ins}(1,4,5)\text{P}_3$ can promote oscillating responses in the membrane potential of *Xenopus* oocytes²¹, suggesting that a persistent supply of $\text{Ins}(1,4,5)\text{P}_3$ could be able to generate repetitive spiking. However only a single transient increase in aequorin luminescence and membrane hyperpolarization was observed when golden hamster eggs were injected with $\text{Ins}(1,4,5)\text{P}_3$, but microinjection of GTP γS evoked decaying repetitive transients (indistinguishable from those evoked by sperm), which were blocked by GDP βS (ref. 13); this result suggests that $[\text{Ca}^{2+}]_i$ spiking is regulated at a point between G-protein activation and $\text{Ins}(1,4,5)\text{P}_3$ formation.

Spikes of $[\text{Ca}^{2+}]_i$ could be generated by relatively small fluctuations in $\text{Ins}(1,4,5)\text{P}_3$ if the release of stored Ca^{2+} were mediated by a highly cooperative mechanism, but until now, studies of $\text{Ins}(1,4,5)\text{P}_3$ binding to its receptor^{22,23} or of $\text{Ins}(1,4,5)\text{P}_3$ -induced Ca^{2+} release^{22,24} show no evidence for cooperativity.

Several mechanisms could be involved in $[\text{Ca}^{2+}]_i$ spike generation in the absence of any fluctuations in $\text{Ins}(1,4,5)\text{P}_3$. (1) Ca^{2+} could synergize with $\text{Ins}(1,4,5)\text{P}_3$ in releasing Ca^{2+} . This is unlikely because Ca^{2+} has little effect on binding of $\text{Ins}(1,4,5)\text{P}_3$

to its receptor in peripheral tissues²² and is actually inhibitory in brain²³. (2) There could be a direct $[\text{Ca}^{2+}]_i$ -induced Ca^{2+} release as observed in heart²⁵. (3) Raised $[\text{Ca}^{2+}]_i$ will cooperatively activate Ca-ATPases^{26} to remove Ca^{2+} from the cytosol. (4) At the same time, although not in response to raised $[\text{Ca}^{2+}]_i$, the rate of release of Ca^{2+} from the store could decrease as the store depletes.

The interspike changes in $[\text{Ca}^{2+}]_i$ are reminiscent of those of the pacemaker potential in rhythmogenic cells of the heart²⁷ or the interspike potential in a motor neuron stimulated by a constant current⁸. The analogy to the pacemaker inward current may be a slowly rising level of $[\text{Ca}^{2+}]_i$, released from internal stores by $\text{Ins}(1,4,5)\text{P}_3$ (which may itself be either slowly rising or constant), until a 'threshold' $[\text{Ca}^{2+}]_i$ is reached to cause a regenerative rise in $[\text{Ca}^{2+}]_i$ by one of the mechanisms mentioned above. Removal of external Ca^{2+} reduces the spiking frequency in endothelial cells, and ultimately the spiking stops. (Similarly, in hamster eggs, microinjected GTP γS evoked only a single hyperpolarization spike in the absence of external Ca^{2+} even though repetitive spiking was seen in the presence of external Ca^{2+} (ref. 13).) Extracellular Ca^{2+} is required to maintain spiking, either by maintaining the intracellular Ca^{2+} store in a nearly full state, or because extracellular Ca^{2+} is part of the source for the 'pacemaker' rise in $[\text{Ca}^{2+}]_i$. Our results indicate that both of these possibilities could be involved.

The spike and plateau response with very high concentrations of agonist could result from overwhelming $\text{Ins}(1,4,5)\text{P}_3$ levels that initially discharge the store and then keep it discharging as fast as it is refilled by Ca^{2+} ; the $[\text{Ca}^{2+}]_i$ level reached would now reflect the balance between refilling the Ca^{2+} store and Ca^{2+} extrusion and resequestration. Removal of either ligand (see Fig. 1) or external Ca^{2+} (data not shown) then causes $[\text{Ca}^{2+}]_i$ to fall rapidly to basal values.

Physiological and pathological implications of a frequency-modulated system. Dose-dependent agonist-evoked $[\text{Ca}^{2+}]_i$ spikes have now been seen in hepatocytes^{6,7} and endothelial cells, and in both cases the results suggest that the frequency of transient rises in $[\text{Ca}^{2+}]_i$ may be the important factor in controlling the cellular response to agonist. A frequency-modulated system may allow for more precise regulation than an amplitude-modulated one, particularly when the effect of Ca^{2+} is highly cooperative, such as when its actions are mediated via calmodulin, as discussed by Cobbold and colleagues⁷. In this case, control by altering a static $[\text{Ca}^{2+}]_i$ level would require a very fine control of $[\text{Ca}^{2+}]_i$ to get a graded response. In contrast, frequency modulation would produce a response which was proportional to the frequency because the frequency would determine the fraction of time for which $[\text{Ca}^{2+}]_i$ was high enough to fully activate the response.

Spiking in post-capillary venule endothelial cells in response to histamine may have important physiological and pathological implications. Low doses of histamine (<10 μM), in the range where we have observed asynchronous spiking in endothelial cell $[\text{Ca}^{2+}]_i$, may be of physiological relevance. Studies of cultured human umbilical vein endothelial cells have shown that histamine in the dose range 0.1–10 μM stimulates release of prostacyclin^{28,29}, platelet activating factor³⁰, Von Willebrand factor⁴ and tissue-type plasminogen activator³¹. These effects would not require the coordinated action of neighbouring cells. Higher doses of histamine (>10 μM), where we observed a maintained elevation in $[\text{Ca}^{2+}]_i$ in all the cells, may be important in pathological conditions. Concentrations of histamine in excess of 10 μM are required both to produce changes in permeability in cultured monolayers of human endothelial cells^{3,32}, and to cause oedema, *in vivo*, in the rat microcirculation³³. Such gross permeability changes are likely to require the coordinated contraction of neighbouring endothelial cells and could therefore be caused by the maintained rise in $[\text{Ca}^{2+}]_i$ in neighbouring cells. Thus, the histamine dose-dependent $[\text{Ca}^{2+}]_i$ spiking which merges into a maintained $[\text{Ca}^{2+}]_i$ rise at high doses of histamine

could fulfil two functions. It could provide a sensitive frequency-modulated control of the release of various agents at low doses although only allowing at high doses the coordinated response necessary to produce oedema.

We thank Dr Roger Tsien for supplying us with his software for image acquisition and analysis, Dr Andrew Davis for help

with implementing the software, Dr Chris Benham for helpful discussion, Messrs Colin Burns, David Harrison and Allan Long for technical support, and Dr Jeremy Pearson for supplying us with endothelial cells.

Note added in proof: Recently a highly cooperative effect of $\text{Ins}(1,4,5)\text{P}_3$ on Ca^{2+} release has been reported³⁷.

Received 1 March; accepted 10 June 1988.

- Schachter, M. (ed.) *Histamine and Antihistamines* Vol. 1 *International Encyclopedia of Pharmacology and Therapeutics* Section 74 (Pergamon, Oxford, 1973).
- Crone, C. C. *News physiol. Sci.* **2**, 16–18 (1987).
- Rotrosen, D. & Gallin, J. I. *J. Cell Biol.* **103**, 2379–2387 (1986).
- Hamilton, K. K. & Sims, P. J. *J. clin. Invest.* **79**, 600–608 (1987).
- Lo, W. W. Y. & Fan, T.-P., D. *Biochem. biophys. Res. Commun.* **148**, 47–53 (1987).
- Woods, N. M., Cuthbertson, K. S. R. & Cobbold, P. H. *Nature* **319**, 600–602 (1986).
- Woods, N. M., Cuthbertson, K. S. R. & Cobbold, P. H. *Cell Calcium* **8**, 79–100 (1987).
- Hille, B. *Ionic channels of excitable membranes* (Sinauer, Sunderland, 1984).
- Schlegel, W., Winiger, B. P., Mollard, P., Vacher, P., Wuarin, F., Zahnd, G. R., Wollheim, C. B., Dufy, B. *Nature* **329**, 719–721 (1987).
- Putney, J. W. Jr *Cell Calcium* **7**, 1–12 (1986).
- Merritt, J. E. & Rink, T. J. *J. biol. Chem.* **262**, 17362–17369 (1987).
- Ambler, S. K., Poenie, M., Tsien, R. Y. & Taylor, P. J. *J. biol. Chem.* **263**, 1952–1959 (1988).
- Miyazaki, S. *J. Cell Biol.* **106**, 345–353 (1988).
- Harden, T. K., Hawkins, P. T., Stephens, L., Boyer, J. L. & Downes, C. P. *Biochem. J.* **252**, 583–593 (1988).
- Orellana, S., Solski, P. A. & Brown, J. H. *J. biol. Chem.* **262**, 1638–1643 (1987).
- Morris, A. J., Downes, C. P., Harden, T. K. & Mitchell, R. H. *Biochem. J.* **248**, 489–493 (1987).
- Biden, T. J. & Wollheim, C. B. *J. biol. Chem.* **261**, 11931–11934 (1986).
- Shears, S. B., Parry, J. B., Tang, K. Y., Kirk, C. J. & Mitchell, R. H. *Biochem. J.* **246**, 139–147 (1987).
- Irvine, R. F. & Moore, R. M. *Biochem. J.* **240**, 917–920 (1986).
- Morris, A. P., Gallacher, D. V., Irvine, R. F. & Petersen, O. F. *Nature* **330**, 653–655 (1987).
- Taylor, C. W., Berridge, M. J., Brown, K. D., Cooke, A. M. & Potter, B. V. L. *Biochem. biophys. Res. Commun.* **150**, 626–632 (1988).
- Spat, A., Bradford, P. G., McKinney, J. S., Rubin, R. P. & Putney, J. W. Jr *Nature* **319**, 514–516 (1986).
- Supattapone, S., Worley, P. F., Baraban, J. M. & Snyder, S. H. *J. biol. Chem.* **263**, 1530–1534 (1988).
- Burgess, G. M. et al. *Nature* **309**, 63–66 (1984).
- Fabiato, A. *J. gen. Physiol.* **85**, 247–289 (1985).
- Lew, V. L., Tsien, R. Y. & Miner, C. *Nature* **298**, 478–481 (1982).
- Noble, D. *The initiation of the heartbeat* 2nd edn (Clarendon, Oxford, 1979).
- Baenziger, N. L., Force, L. E. & Becherer, P. R. *Biochem. biophys. Res. Commun.* **92**, 1435–1440 (1980).
- Resink, T. J., Gregorian, G. Y., Moldabaeva, A. K., Danilov, S. M. & Bühler, F. *Biochem. biophys. Res. Commun.* **144**, 438–446 (1987).
- McIntyre, T. M., Zimmerman, G. A., Satoh, K. & Prescott, S. M. *J. clin. Invest.* **76**, 271–280 (1985).
- Hanss, M. & Collen, D. *J. lab. clin. Med.* **109**, 97–104 (1987).
- Killackey, J. J. F., Johnston, M. G. & Movat, H. Z. *Am. J. Pathol.* **122**, 50–61 (1986).
- Guth, P. H. & Hirabayashi, K. *Microsc. Res.* **25**, 322–332 (1983).
- Hallam, T. J., Pearson, J. D. & Needham, L. A. *Biochem. J.* **255**, 243–249 (1988).
- Grynkiewicz, G., Poenie, M. & Tsien, R. Y. *J. biol. Chem.* **260**, 3440–3445 (1985).
- Tsien, R. Y. & Poenie, M. *Trends biol. Sci.* **11**, 450–455 (1986).
- Meyer, T., Holowka, D. & Stryer, L. *Science* **240**, 653–656 (1988).

Identification of predictive sequence motifs limited by protein structure data base size

Marianne J. Rومان* & Shoshana J. Wodak†

* Plant Genetic Systems and † Unité Conformation de Macromolécules Biologiques, Université Libre de Bruxelles, CP 160, P2, Ave. P. Héger, 1050 Bruxelles, Belgium

Associations between short amino acid sequence patterns and protein secondary structure classes can be found by searching a data base of known protein structures. Analysis of these associations suggests that secondary structure of proteins can be determined locally by sequence motifs of high predictive value, but at present our ability to find these motifs is limited by the size of the available data bases.

DERIVING the three-dimensional structure of a protein from its amino acid sequence remains a major challenge in molecular biology that available methods have so far been unable to meet in a satisfactory way.

Secondary structure predictions, whether based on statistical analyses of the data bases^{1–4}, or on folding schemes and physical or chemical considerations⁵, are very widely used, but yield predictions that are correct only up to about 60%⁶.

Different methods using rules derived from a detailed analysis of experimentally determined three-dimensional protein structures have recently offered interesting perspectives. These include the assignment of patterns in amino acid sequence to local structural motifs^{7–9}, as well as higher order rules based on observed topologies of the polypeptide chain^{10,11}. However, these methods have often been applied only to subsets of protein structures or to specific structural motifs and shown to produce limited improvements within these subsets.

The reasons underlying the low success score of structure prediction methods are unclear. Is it due to structural information being so strongly dependent on context and interactions between residues that are distant along the chain? Or is the data base of known protein structures too small to produce valid rules? Is it possible to find patterns that characterize secondary structures well and, if so, what is special about them?

To address these questions we take the rule-based approach

to protein structure prediction at its foundations. A data base containing 75 proteins of known structure is searched with simple sequence patterns of the form Ala-X-X-Phe-X-Glu, looking for those associations of patterns with four classes of secondary structure which are sufficiently good to satisfy a certain criterion. Finding many such associations, we then assess their predictive value using an internal test based on dividing the data base into 'learning' and 'test' subsets¹².

Our results show that highly predictive patterns do occur, suggesting that local primary sequence can determine secondary structure independently of tertiary interactions. But at present the size of the available data bases limits our ability to find such patterns, because they need to appear frequently enough in the database in order to weed out poorly predictive patterns. Implications of our findings for future ability to predict protein structure from sequence are discussed.

Pattern language

To allow a convenient representation of all possible amino acid sequence information and secondary structure assignments along the polypeptide chain, we developed a general one-dimensional pattern language.

Consider a finite sequence whose elements belong to the finite set $\{x_i\}$. Here the elements x_i are either amino acid residues or secondary structure assignments.

A subclass of patterns along that sequence (called elementary patterns) can be written as:

$$E(x) = n_0x_0 \cdots \& n_i x_i \cdots \& n_f x_f \quad (1)$$

where $n_0 \leq \cdots \leq n_i \leq \cdots \leq n_f$ are integers representing relative shifts in position along the sequence, and the symbol '&' denotes the logical 'and' operator. The length L of the pattern $E(x)$ is defined as $L = (n_f - n_0 + 1)$. With x_i representing amino acid residues and using the one letter code, the pattern (0A&1C&3D) is of length $L = 4$ and represents the amino acid sequence Ala-Cys-X-Asp, where X is an arbitrary residue.

The most general patterns $P(x)$ can be expressed as disjunctions of elementary patterns $E_i(x)$, using the logical 'or' operator denoted as '|':

$$P(x) = E_0 \cdots | E_i(x) \cdots | E_q(x) \quad (2)$$

An example of such general patterns is (0A&1C&3D)|(1E&2F), which translated into common notation means that over a stretch of 4 residues along the primary sequence we have the following amino acids: Ala-Cys-X-Asp or X-Glu-Phe-X.

The connectives '&', '|', and 'shift' are equivalent to the 'and', 'or' and 'next' (and 'previous', for negative shifts) of first order temporal logic¹³. The negation ('not') and the connectives 'sometimes' and 'always' which are also part of the temporal logic formalism, are not necessary in our case because of the finite number of values taken up by x_i and the finite length of the sequence.

Sequence and secondary structure

In this study we generate patterns in amino acid sequence and look for those that characterize specific secondary structures in a data base of 75 known protein structures (listed in Fig. 1) representing a total of 13,214 amino acid residues. These structures have been selected from the Brookhaven data bank¹⁴. They have a resolution of 2.5 Å or better, have been refined crystallographically and exhibit less than 50% sequence homology.

The assignment of secondary structure is computed from the published coordinates using the program DSSP¹⁵. Four classes of secondary structure are considered: H (3/10, α and π -helix conformations); E (β -strand); T (turns or bends); C (coil).

The secondary structure assignment is made on a residue basis and can therefore also be expressed as a one-dimensional pattern using our pattern language. We thus look for relations of the type:

$$P(R) \Rightarrow P'(S) \quad (3)$$

where R denotes amino acid residues and S secondary structure assignments. This relation states that a given pattern in amino acid sequence is unambiguously associated with a pattern in secondary structure. This requirement is clearly too restrictive. Condition (3) is therefore relaxed by allowing an error margin:

$$P(R)(M_+, M_-) \Rightarrow P'(S) \quad (4)$$

where M_+ and M_- are respectively the number of times relation (4) is true (the number of matches), and the number of times the relation is false (the number of mismatches). The error margin is expressed as the ratio of mismatches to the total number of times the pattern occurs.

As the first step in our approach, this study considers only elementary patterns in amino acid sequence, corresponding to expression (1). These represent the occurrence of specific amino acids at distinct positions along the primary sequence. The analysis of patterns implying the occurrence of different residues at the same sequence position, corresponding to a consensus sequence or to physical properties (hydrophobic, bulky, polar and so on), and expressed in our formalism using expression (2), will be described elsewhere. In addition, in a given sequence pattern, at most three residues are specified over the pattern

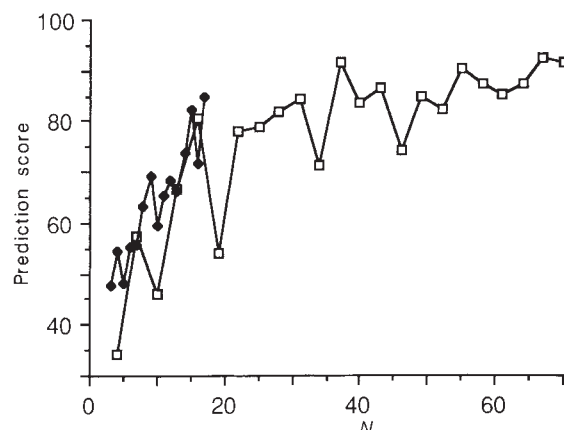


Fig. 1 Prediction score of amino acid patterns that characterize secondary structures as a function of N , the number of times they crop up in the data base. The prediction score is expressed as the percentage of the hits which yield correct predictions in proteins outside the learning set. $\square$, Results obtained for 2-residue patterns; $\blacklozenge$, results obtained for 3-residue patterns. For 2-residue patterns, points plotted represent data combined from 3 successive values of N ($N-1$, N , $N+1$). The number of patterns contributing to each N is $Q = \sum_{i=1}^{75} q_i$, where i are the learning sets, each having one test protein, and q_i the number of patterns that occur N times in an individual learning set and that have a hit in the test protein. Some examples of couples of Q and N values: for 2-residue patterns (Q , N) = (138, 3), (76, 15), (129, 25), (68, 40), (11, 57), (0, 70), for 3-residue patterns (Q , N) = (29,556, 4), (1,239, 10), (20, 17). Patterns lengths span the range of 1-9 residues. The following proteins given by their Protein Databank code¹⁴ are used in the present study: 1APR, 2ACT, 1ACX, 4ADH, 5CHA, 2ABX, 2ALP, 1PPT, 2AZA, 1TPP, 3ICB, 2CAB, 5CPA, 8CAT, 2MT2, 2CTS, 2CNA, 1CRN, 2SOD, 2B5C, 2CCY, 1CCR, 2CYP, 3C2C, 2CDV, 1CY3, 1CC5, 155C, 451C, 3DFR, 2EST, 4APE, 2FD1, 3FXC, 4FXN, 1GCR, 2GN5, 1GPI, 1HMZ, 1ECD, 4HHB, 2LHB, 1HIP, 1FB4, 1INS, 2PKA, 1CTF, 4LDH, 2LH4, 1LZM, 1LZ1, 2MDH, 1MLT, 1MBD, 1NXB, 9PAP, 1CPV, 1BP2, 1PCY, 2PAB, 2SGA, 3SGB, 3RP2, 1RHD, 1RN3, 4RXN, 2STV, 1SN3, 2SNS, 1SBT, 3TLN, 1TIM, 5PTI, 3WGA, 2YHX.

interval. It would not make sense to specify more than three residues because the corresponding patterns do not occur often enough in our data base. Finally, we limit the total length of sequence patterns to nine residues ($0 \leq n_f - n_0 \leq 8$), roughly corresponding to two turns of a helix, a choice justified *a posteriori* by our analysis. Thus in total, 37 pattern types are considered: x ; $x\&my$ with $1 \leq m \leq 8$; $x\&my\&nz$ with $1 \leq m < n \leq 8$, where x , y , z are any of the 20 naturally occurring amino acid residues.

In patterns of secondary structure we distinguish two types. Let us consider sequence regions of length L defined by the matches of a given amino acid pattern in the database. We require that within each of these regions a specific secondary structure must occur either at a given position—the same in all cases—or at least m times at any position in the region. The former are termed 'position patterns' and the latter 'interval patterns'. In interval patterns, $m \geq 1$ for patterns of length $L \leq 4$ and $m \geq L/2$ for patterns of length $L > 4$.

The data base of 75 proteins is systematically scanned for the occurrence of all sequence patterns that belong to the 37 types described above, using an automatic pattern matching procedure coded in FORTRAN and C languages. Patterns that satisfy relation (4) are retained using the following protocol: the number N of occurrences of a given pattern, defined as the sum of its matches and mismatches ($N = M_+ + M_-$), must be at least three; the maximum allowed error M_-/N is either ≤ 0.2 when the characterized secondary structure is one of the four classes mentioned above, or ≤ 0.1 when the secondary structure is T|C, a fifth redundant class that occurs about twice as often as the

other four classes¹⁶. Note that we do not retain a pattern as characterizing (T|C) if it already characterizes either T or C separately, unless the (T|C) pattern has a lower error score.

With this protocol, a total of 22,443 different sequence patterns, associated by expression (4) to 58,964 secondary structure patterns, have been retained (we use the convention of considering as two different patterns those of the type $nx&my$, where x and y denote secondary structure assignments). The procedure took 51 minutes of central processing unit time on a SUN 3/50 workstation.

We find that the fraction of patterns characterizing each of the five types of secondary structures roughly corresponds to the relative frequencies of the secondary structures in the learning set. Patterns with one specified residue have not been retained as they do not satisfy relation (4), implying that single amino acids¹ are not sufficient to characterize secondary structure.

Predictive power of amino acid patterns

To assess the predictive power of retained amino acid patterns, they are matched against sequences of proteins whose structure is known. For each match, the secondary structure assignments (the 'observed structure') and secondary structure pattern (the 'predicted structure') are compared. When this procedure is applied to all 75 proteins in the data base, the predicted secondary structure is identical to the observed one in 96% of the cases. This is not very surprising if we consider that these proteins belong to the very same set from which the patterns have been derived.

The only valid test of the predictive power of our patterns is to apply the same procedure to proteins outside the learning set which are moreover not homologous to members of this set. This is done by a classical cross-validation procedure¹²: proteins are removed one at a time from the original set of 75 structures. Patterns are freshly compiled on the remaining reduced set and then tested in the removed protein. The result is a catastrophic drop in prediction score. It is now 51%, compared to 96% previously obtained on members of the learning set. This score can be raised to 64% if we include patterns that are 'nearly correct', in which the position of the predicted secondary structure is displaced by one residue from the expected one (for position patterns), or the predicted secondary structure appears $m-1$ times in the interval instead of the expected m times, with $m > 1$ (for interval patterns).

Thus, we encounter the same barrier of about 60% success rate as with other available methods. This result leads to the first important conclusion of our study: the low accuracy of structure prediction methods is not due to a lack of generality or the non-systematic nature of the approach, as ours is as systematic as can be. Other more fundamental reasons must therefore be considered. Two have often been invoked but not rigorously proven^{3,6,18}: (1) the learning set of known protein structures may be too small to produce valid prediction rules and (2) crucial information on contributions from residues that are far apart in the sequence but close in the three-dimensional structure are not adequately represented in the prediction methods.

To evaluate these hypotheses, we ask these two questions. Do subclasses of retained amino acid patterns have a better than average predictive power? How does the predictive power of patterns vary with the number of proteins in the learning set?

We find that position and interval patterns are, on average, equally predictive. Similarly, no difference in the predictive power of patterns can be detected when the statistics are segregated into the five types of secondary structures. The scores obtained are respectively 41% for (E); 47% for (H); 43% for (T); 41% for (C) and 60% for (T|C). The higher score obtained for (T|C) patterns simply reflects the stricter error limit used to define them (only 10%), as well as the fact that turn and coil put together occur about twice as often as the other types of

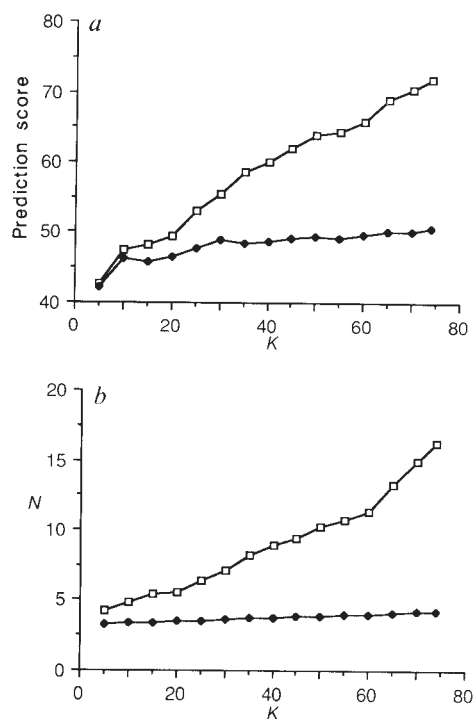


Fig. 2 Properties of patterns that characterize secondary structures, as a function of K , the number of proteins in the data base, a , Prediction score of 2-residue patterns ($\square$) and 3-residue patterns ($\blacklozenge$) as a function of K . b , Average number of occurrences (N) of 2-residue patterns ($\square$) and 3-residue patterns ($\blacklozenge$) as a function of K . $\langle N \rangle = (\sum_{i=1}^M N_i) / M$, with N_i being the number of times an individual (retained) pattern i crops up in the data base and M is the total number of retained patterns in the data base. Pattern lengths span the range of 1-9 residues.

secondary structure¹⁶. This is of course consistent with earlier findings that these regions can be predicted with higher accuracy¹⁷.

Next, the influence of the number of specified residues in the pattern is analysed. We find that the average prediction score is higher for patterns with two specified residues (72%) than for patterns with three specified residues (51%). Figure 1 shows that this is only due to the higher number of occurrences of 2-residue patterns compared to 3-residue patterns in the learning set, and not to an intrinsic property of the patterns. Indeed, at an equal number of occurrences quite the opposite is true: the predictive power of 3-residue patterns is distinctly superior to that of 2-residue patterns. The shape of the curves in Fig. 1 suggests that this advantage should be conserved when the prediction score of the respective pattern types reach their asymptotic values. This observation and our finding that 1-residue patterns lack predictive power, suggest that n -residue patterns are more predictive than $(n-1)$ -residue patterns (for an equal number of occurrences), in agreement with the valid assumption that the former should have a higher information content.

Another important point illustrated in Fig. 1 is that the prediction score of patterns increases steadily with N , the number of times they occur in the learning set. We find that, on the average, patterns that occur three times in the learning set are correct only about 40% of the time, whereas those that occur 60 times are correct 85-90% of the time, a simple consequence of more reliable statistics. The minimum value of N required to retain a pattern in the learning phase should therefore be much higher than three (used here) if we want to have predictive patterns. To attain an average prediction score of about 80% it should be set to about 15. Unfortunately, only 123 patterns out of a

Table 1 Examples of patterns with high prediction score in the different classes of secondary structure

<i>L</i>	<i>S</i>	<i>N</i>	<i>I/P</i>	s-s	Amino acid patterns	(<i>M</i> ₊ , <i>M</i> ₋)	Secondary structure patterns	Score (%)
4	3	13	P	E	G&1G&3L	(12, 1)	2E&3E	80
4	3	12	I	E	I&2G&3G	(10, 2)	[4!1E]	80
5	3	16	I	H	A&1A&4K	(14, 2)	[5!3H]	87
4	2	13	P	H	L&3C	(11, 2)	0H	89
					L&3C	(12, 1)	1H	
3	3	14	P	T	G&1D&2S	(13, 1)	0T	87
3	3	14	I	T	D&1S&2G	(12, 2)	[3!1T]	85
					G&1D&2S	(14, 0)	[3!1T]	
					G&1S&2T	(12, 2)	[3!1; 2T]	
4	3	15	I	C	S&2G&3S	(13, 2)	[4!1; 2C]	91
4	3	12	I	C	S&1G&3P	(12, 0)	[4!1; 3C]	87
					S&1T&3A	(11, 1)	[4!1C]	
4	2	31	I	T/C	S&3C	(28, 3)	[4!1(T/C)]	90
5	3	13	P	T/C	S&1G&4V	(12, 1)	1(T/C)	92
					G&3T&4V	(12, 1)	0(T/C)	
					G&3V&4T	(12, 1)	0(T/C)	

Columns 1–5 give the specifications of the pattern categories: *L* designates the pattern length, *S* the number of residues specified in the pattern, *N* the total number of occurrences in the data base. *I* and *P* stand for interval and position patterns respectively, and s-s for secondary structure (see text). Column 6 lists all the amino acid sequence patterns in each category, retained from the full learning set. Column 7 lists the number of matches (*M*₊) and mismatches (*M*₋) of each specific pattern and column 8 the predicted secondary structure patterns. The notation used for position patterns is explained in the text. To express interval patterns in a simple way we introduce the following redundant notations: [*L**x*] represents a sequence interval of length *L* over which *x* holds *L* times, [*L*!*m**x*] represents an interval of length *L* in which *x* occurs at least *m* times and [*L*!*m*; *n**x*] represents an interval of length *L* in which *x* occurs at least *m* times and at most *n* times. The rightmost column lists the average prediction score defined as the fraction of correct predictions by patterns of the given category, obtained from learning sets excluding one protein at a time. Note that the higher than 80% score obtained for patterns with *N* < 15 in specific categories is not inconsistent with the fact that a broader class of patterns (counting 123 members) specified only by the requirement that *N* ≥ 15 (see text), has a uniform (average) score of 80%, independently of the values of other parameters (*L*, *S*, *I/P*, s-s).

total of 58,964 occur at least 15 times in our learning set. Ninety-seven of them are 2-residue patterns and 26 are 3-residue patterns.

We now examine the influence of the number of proteins in the learning set on the predictive power of our patterns. This is done by performing predictions on each of the 75 proteins, taken one at a time, using patterns compiled from learning sets that contain an increasing number of proteins (the first 5, first 10, first 15 proteins, and so on, up to 74, of our original set) but that exclude the protein on which the predictions are made.

We find (Fig. 2*a,b*) that the fraction of correct patterns (or the prediction score) increases with the number of proteins in the learning set, and see that the crucial parameter determining this fraction is, once more, the average number of occurrences *N* of the corresponding patterns in the data base. The prediction score of 3-residue patterns—unlike that of 2-residue patterns—appears to be nearly insensitive to the data base size. This simply reflects the fact that on average these patterns occur less than three times over the entire range of data base sizes (1.6 times at the maximum data base size of 74 proteins).

Only two parameters are found so far to have a crucial influence on the prediction score of our patterns: the number of specified residues in the pattern and the number of times a pattern occurs in the data base. What about pattern length? Are longer patterns more predictive than shorter ones? Within the limit of our analysis, the answer is no. We find that the percentage of correct patterns decreases smoothly from 63% for patterns of length *L* = 3 residues to 48% for *L* = 9. A more detailed analysis shows that this is again related to pattern occurrence *N*. For *L* = 3, 13% of all retained patterns occur 15 times or more in the data base, while for *L* = 9 this fraction is only 0.3%; yet, the average prediction score of patterns that occur at least 15 times is about 80%, irrespective of pattern length. These results imply that over interval lengths of 1–9 residues, and with two or three specified positions, patterns that best characterize

secondary structure are short (*L* < 5), largely justifying our original decision not to consider patterns longer than nine residues. The probability that two or even three residues positioned in a pattern have an influence on its structure is indeed higher in short patterns than in long ones, where more unspecified residues can easily override this influence.

A sample of compiled patterns that characterize the four classes of secondary structure well is given in Table 1. Although these patterns contain some amino acids that have previously been associated with the corresponding secondary structures^{1–3}, the patterns themselves are new and should therefore be helpful in improving currently available prediction methods (manuscript in preparation).

Discussion

One of the major conclusions of our study is that short amino acid patterns do exist which characterize secondary structures in proteins independently of tertiary interactions.

This is not inconsistent with the finding that short segments of identical amino acid sequence can adopt different conformations in proteins¹⁸. Here, we retain patterns that uniquely characterize secondary structures with a specified error margin and reject those which do not. Therefore, our patterns match only regions of the polypeptide sequence whose conformation is locally determined. Alternatively, as our patterns contain gaps, they may overlap short structurally variable sequences, and therefore yield, at least in principle, correct structure predictions for these sequences.

A second important point of the present analysis is the demonstration that the information content of the data base determines—more than any other factor—our ability to link unambiguously amino acid patterns to structure. But then, how does it affect our capacity to make predictions?

Presently, 28% of a protein sequence at most can be predicted with 80% accuracy by the subset of 123 most frequent patterns,

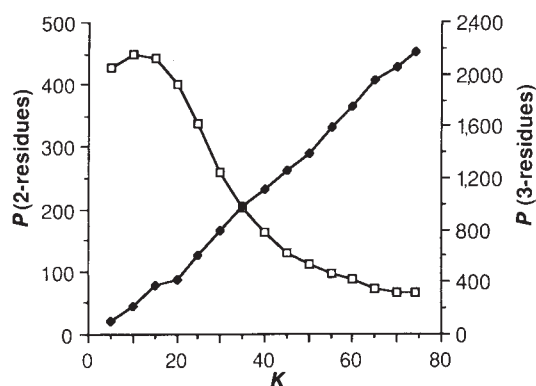


Fig. 3 Average number P of predictions per protein, provided by patterns that characterize secondary structures, as a function of K , the number of proteins in the database. P is equal to the average number of hits in a test protein of the patterns (of length 1–9 residues) compiled from a database containing K proteins. At the maximum value of $K = 74$, 2-residue patterns ($\square$) provide 66 predictions per protein whose average length is 180 residues. If we assume that each pattern match predicts the conformation of one residue, only 37% of a protein sequence will be predicted with 72% accuracy (Fig. 2a). P for 3-residue patterns ($\blacklozenge$) increases steadily from 94 for $K = 5$ to 2,168 for $K = 74$. This increase resembles that observed for 2-residue patterns in the region where $K < 10$. It is due in both cases to the average number of pattern occurrences being less than three over the respective ranges of database size. By extrapolation it is therefore tempting to assume that the 3-residue curve will continue to rise until about $K = 130$ (where an average of three occurrences will be reached) with an expected maximum value of P or more than 3,000, then start to decrease, and finally level off like the 2-residue curve. We speculate that this is likely to occur at K values around 1,500 ($= 75 \times 20$), where the projected value of P could be as high as 450, or about seven times the P value provided by 2-residue patterns today. This will correspond to an average of about 2.5 accurate predictions per residue instead of the present meagre 0.4 predictions per residue.

as obtained from the simple computation: $(123 \times 30 / 13,214 = 0.28)$, where 30 is the average number of occurrences of the 123 patterns in the 13,214 residue learning set. This leaves at least 70% of the protein sequence unpredicted. To avoid having unpredicted regions it is therefore necessary to consider less accurate patterns, causing a drop in the score to below 60%.

Our analysis can also be used to estimate future progress in structure prediction, which will crucially depend on availability of a larger number of accurate patterns that match an increased fraction of the protein residues. Clearly however, such estimates remain highly speculative at present.

Figures 1 and 2 suggest that pattern accuracy will continue to increase with data base size, reaching a score of 85–90% for 2-residue patterns, and possibly a higher score for 3-residue patterns. We estimate that the average prediction score of 3-residue patterns will reach the present score of 2-residue patterns when the learning set is 20 times the present size (containing 1,500 proteins with an average length of 180 residues), since 3-residue patterns will occur as frequently then as 2-residue patterns do now.

The average number of predictions per protein P , which is a measure of whether a given residue is likely to have a prediction but not of how good the prediction will be, is given as a function of database size in Fig. 3. Below a certain size when the average number of pattern occurrences is less than three, P increases. Above three occurrences, P for 2-residue patterns—the only kind that occurs often enough in the data base—declines with increasing K (the number of proteins in the data base) as the poor predictive patterns are weeded out. This decline appears to level off at the maximum size of the learning set (74 proteins). Therefore, larger data bases should not be expected to yield

more predictions by 2-residue patterns than we already have now. They could provide the same number of predictions with the present accuracy or fewer predictions with higher accuracy, depending on pattern selection criteria.

Our data provides fewer clues on the expected contribution from 3-residue patterns in larger data bases. Yet if we assume that P for 3-residue patterns will display over the range of 75–1,500 proteins the same behaviour as P for 2-residue patterns over the range of 5–75 proteins, then a data base 20 times the present size should yield many more 3-residue patterns. This will allow us to predict with high accuracy a much larger fraction of a protein sequence than at present (see Fig. 3). This, together with contributions from patterns having four or more specified residues, and with model-building methods that optimize tertiary interactions, may yield satisfactory structure predictions.

Therefore, utmost priority should be given to scaling up efforts in protein structure determination. Structure prediction, by the approach described here, or by exploiting sequence homology¹⁹, as well as structure determination by X-ray diffraction²⁰ and nuclear magnetic resonance²¹, will increasingly rely on available structural information. Also, as the size of protein data bases grows, methods of automatic extraction and verification of patterns, such as those described here, will become indispensable, and have the potential for many future applications.

We are grateful to J. Rodriguez, J. Richelle, M. Claessens, M. Huysmans and M. Moreau for helpful discussions and use of the protein data base. We also thank B. Marchal for exchange of ideas in the initial stages of this study and acknowledge l'Institut pour l'Encouragement de la Recherche Scientifique dans l'Industrie et l'Agriculture for financial support.

Received 28 March; accepted 27 June 1988.

1. Chou, P. Y. & Fasman, G. D. *Biochemistry* **13**, 222–245 (1974).
2. Garnier, J., Osguthorpe, D. J. & Robson, B. *J. molec. Biol.* **120**, 97–120 (1978).
3. Gibrat, J.-F., Garnier, J. & Robson, B. *J. molec. Biol.* **198**, 425–443 (1987).
4. Finkelstein, A. V. & Ptitsyn, O. B. *J. molec. Biol.* **62**, 613–624 (1971).
5. Lim, V. I. *J. molec. Biol.* **88**, 873–894 (1974).
6. Kabsch, W. & Sander, C. *FEBS Lett.* **155**, 179–182 (1983).
7. Sibanda, B. L. & Thornton, J. M. *Nature* **316**, 170–174 (1985).
8. Edwards, M. S., Sternberg, M. J. E. & Thornton, J. M. *Protein Engineering* **1**, 173–181 (1987).
9. Wierenga, R. K., De Maeyer, M. C. H. & Hol, W. G. J. *Biochemistry* **24**, 1346–1357 (1985).
10. Taylor, W. R. & Thornton, J. M. *J. molec. Biol.* **173**, 487–514 (1984).

11. Cohen, F. E., Abarbanel, R. M., Kuntz, I. D. & Fletterick, R. J. *Biochemistry* **25**, 266–275 (1986).
12. Stone, M. J. *R. Statist. Soc. Series B* **36**, 111–147 (1974).
13. Rescher, N. & Urquhart, A. *Temporal Logic* (Springer, Berlin, 1971).
14. Bernstein, F. C. *et al. J. molec. Biol.* **112**, 535–542 (1977).
15. Kabsch, W. & Sander, C. *Biopolymers* **22**, 2577–2637 (1983).
16. Crawford, J. L., Lipscomb, M. N. & Schellman, C. G. *Proc. natn. Acad. Sci. U.S.A.* **70**, 538–542 (1973).
17. Cohen, F. E., Abarbanel, R. M., Kuntz, I. D. & Fletterick, R. J. *Biochemistry* **22**, 4894–4904 (1983).
18. Kabsch, W. & Sander, C. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1075–1078 (1984).
19. Blundell, T. L., Sibanda, B. L., Sternberg, M. J. E. & Thornton, J. M. *Nature* **326**, 347–352 (1987).
20. Jones, T. A. & Thirup, T. *EMBO J.* **5**, 819–822 (1986).
21. Kraulis, P. J. & Jones, T. A. *Proteins* **2**, 188–201 (1987).

On the variability of the X-ray emission from supernova 1987A

A. C. Fabian & M. J. Rees

Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK

The hard-X-ray emission observed above 10 keV from SN1987A by the X-ray astronomy satellite *Ginga*¹ and by the KVANT module² on the MIR station is reasonably well explained^{3,4} by Compton down-scattered photons from the decay of ⁵⁶Co. The rising X-ray spectrum detected by *Ginga*¹ below 10 keV is not understood so easily: some have suggested⁵ that it is due to the interaction of the expanding ejecta with circumstellar matter, others⁶ explain it as synchrotron radiation from the region immediately around a newly formed pulsar. Neither of these models satisfactorily accounts for both the variability observed in the X-ray emission^{7,8} (in particular a drop of 30% in 1.5 days in January 1988) and the presence of an emission line at ~6.8 keV. We suggest here that the X-ray emission is due to the young neutron star being cocooned in matter from a binary companion. The remnant must therefore contain large holes which can explain the early detection of gamma rays and hard X rays without requiring that the Ni (which decayed into Co) was mixed throughout the ejecta.

The evidence for decreases in the X-ray emission below 10-keV and for a 6.8 ± 0.2 -keV emission line together indicates that most of the X-rays are thermal and from a small region. If this is a blob of dense gas that has just been overtaken by the supernova blast wave, then it will be pressurized and shocked at a speed $\sim (P/\rho)^{1/2}$, where P is the pressure of the blast wave and ρ is the density of the blob. A drop in emission is presumably caused by the blob cooling (although the spectra⁸ indicate more of a spectral hardening). The bremsstrahlung cooling time $t_c \approx 2 \times 10^{11} T^{1/2}/n$ seconds, which implies $n > 10^{10} \text{ cm}^{-3}$ if the gas is to cool radiatively in less than 2 days from $T \approx 10^8 \text{ K}$. Such a shock temperature is reasonable if the full ram pressure is used. The observed X-ray luminosity of $\sim 10^{38} \text{ erg s}^{-1}$ implies a blob size of $< 3 \times 10^{13} \text{ cm}$, which is very much less than the size of the blast wave. Shocked gas could also cool adiabatically, but an expansion timescale of only 1–2 days would imply a similar upper limit on the blob size. The issue then becomes a question as to why such small dense blobs ($< 10^{-2}$ of the size of the blast wave) should be so few that we can see the decay of individual blobs (see also ref. 9), and the existence of any blobs with a density about one million times higher than the expected circumstellar value has no ready explanation.

One obvious place for a small emission region is the centre of the remnant. This is where Bandiera *et al.*⁶ propose that a synchrotron nebula has formed around a rapidly spinning pulsar. Such a model cannot, however, explain the ~6.8-keV emission line which is presumably due to Fe XXV and Fe XXVI. Fluorescent emission occurs at 6.4 keV and is therefore excluded. Doppler shifts in the expanding remnant are too small to shift the line appreciably. (Fluorescent X-rays from cobalt would have the correct energy but most of the Co should now have decayed into Fe.) Also, it is not clear how variations would be produced, although the remnant may be full of holes (perhaps some induced by the pressure of the relativistic fluid from the pulsar) and the pulsar may have a transverse velocity imparted by the explosion so occultations by the remnant may occur.

The ~6.8-keV line indicates that much of the emission is thermal and is not direct synchrotron radiation. Of course, synchrotron radiation may be the source of heat. Low-frequency radiation and relativistic plasma from the pulsar will deposit energy in the innermost expanding ejecta. We estimate that the

fractional size, x , of the pulsar-driven cavity in a constant density remnant is given by $x \approx (\dot{E}_t/E_{\text{SN}})^{1/5}$, where $\dot{E}_t$ is the total pulsar energy trapped in the cavity and E_{SN} is the kinetic energy of the supernova ejecta. Even optimistically, the total pulsar emission is $\sim 10^{-4}$ of the total kinetic energy of the remnant, so that the cavity is now $\leq 10^{15} \text{ cm}$. Putting all the pulsar energy into the mass of gas that was within the cavity (x^3 times about $5 M_\odot$) gives a gas temperature of only $3 \times 10^7 \text{ K}$. In practice the shock temperature will be lower than this, as much of the energy will be kinetic. The shock temperature of any embedded denser blobs will be still lower. Such a model is therefore unable to account for the X-ray emission. Inoue *et al.*⁸ report that the flare has a spectrum characterized by a temperature exceeding 10^8 K .

A more realistic possibility occurs if the synchrotron emission from the pulsar photoionizes filaments (as in the Crab nebula) so that iron is very highly ionized and radiates lines around 6.8 keV. This requires an ionization parameter $\xi = L/nR^2 > 10^3$ (ref. 10), so that the gas density $n < 10^7 L_{38}/R_{14}^2 \text{ cm}^{-3}$, where the ionizing luminosity (essentially in the range 2–20 keV) is in units of $10^{38} \text{ erg s}^{-1}$, and the radius at which the gas lies is in units of 10^{14} cm . The strength of the line emission⁸ means that $n > 1.6 \times 10^9 R_{14}^{-3/2} (fA)^{-1/2}$. f is the volume-filling fraction of the gas and A is its abundance relative to solar values. Such a model then requires $R = 4 \times 10^9 L_{38}^2 Af \text{ cm}$. It can therefore work only if the region is very small, if the ionizing flux is considerably higher than that seen by *Ginga* ($\sim 10^{40}$ rather than $\sim 10^{38} \text{ erg s}^{-1}$ and so most of the emission is occulted from view), and/or if $Af \gg 1$. For the last option, the apparent solar abundance of the iron on a thermal interpretation requires a coincidentally high iron abundance for the matter in the filaments. This may not be unreasonable in the centre of the remnant. The model also requires that the covering fraction of the filaments is high so that there is little escaping radiation which produces fluorescent X-ray line emission from cold gas. This may be a serious problem but is geometry-dependent. Some form of wind or jet from the pulsar may be more efficient in producing X-rays (compare SS433), but a source of matter for the wind is required.

Perhaps the simplest explanation is that the photoionized region is small ($R < 10^{10} \text{ cm}$). The presence of any gas at that radius is then the problem. However fresh matter could originate from a binary companion to the progenitor star. (Models have already been put forward in which SN1987A occurred in a binary system^{11,12}.) Whether the companion or the supernova progenitor is SK-69 203 is not important here, but we note that ~50% of all stars are found to be at least binary so it is not unreasonable to suppose that SN1987A occurred in a binary system. We require only that the companion is fairly close and is losing mass either through a wind or through tidal overflow, perhaps enhanced as the star readjusts after the impact of the explosion. The object is thus a cocooned pulsar resembling Cyg X-3 which has iron-line emission at 6.76 keV, consistent with the energy of the emission line of SN1987A. If the matter is gravitationally attracted to the pulsar then the Eddington limit will apply, agreeing with the observed X-ray luminosity. The binary companion would have helped to make the supernova ejecta asymmetric so that we are now able to see into the core of the remnant. A further consequence is that the cobalt produced from the nickel just above the surface of the neutron star in the explosion is available for viewing now, without requiring mixing to the outside (a velocity increase by several thousand km s^{-1}). The Doppler shift of the 847-keV line from Co decay (ref. 13; W. G. Sandie *et al.*, personal communication) should resolve this.

In this binary model, which can also help to explain some of the other features of the supernova, we predict that the X-ray emission will remain at roughly the same level ($\sim 10^{38} \text{ erg s}^{-1}$) for a long time, with flares, dips and eclipses possible (like Cyg X-3, although we expect a longer orbital period, see ref. 11). If fresh matter continues to swathe the pulsar, then it will not be

directly visible, except perhaps to ultra-high-energy gamma rays. At the present time the total emission from the binary (mostly X rays) is less than one per cent of the total of the luminosity of SN1987A, so the binary need not be detectable at longer wavelengths. On the other hand, a pulsar more powerful than the observed X-ray emission, which is photoionizing a larger region and yet is occulted from our line of sight, will soon be detectable in the total light curve of the remnant. Further X-ray emission may of course originate from the interaction of the blast wave with circumstellar matter and from Co decay. In our preferred model, the cocooned pulsar and its companion will eventually dominate as the remnant becomes optically thin and transparent.

We acknowledge helpful discussion of unpublished Ginga data with Professor Tanaka and his colleagues.

Received 19 May; accepted 13 July 1988.

1. Dotani, T. *et al. Nature* **330**, 230-232 (1987).
2. Sunyaev, R. *et al. Nature* **330**, 227-229 (1987).
3. McCray, R., Shull, J. M. & Sutherland, P. *Astrophys. J.* **317**, L73-L77 (1987).
4. Ebisuzaki, T. & Shibazaki, N. *Astrophys. J.* (in the press).
5. Masai, K., Hayakawa, S., Itoh, H. & Nomoto, K. *Nature* **330**, 235-236 (1987).
6. Bandiera, R., Pacini, F. & Salvati, M. *Nature* (in the press).
7. Tanaka, Y. *Proc. IAU Coll.* **108** (in the press).
8. Tanaka, Y. *Proc. Conf. Compact Objects*, Tokyo (in the press).
9. Chevalier, R. A. *Nature* **232**, 514-516 (1988).
10. Hatchett, S., Buff, J. & McCray, R. *Astrophys. J.* **206**, 847-860 (1976).
11. Fabian, A. C., Rees, M. J., Van den Heuvel, E. P. J. & Van Paradijs, J. *Nature* **328**, 323-324 (1987).
12. Joss, P. C., Podsiadlowski, P., Hsu, J. J. L. & Rappaport, S. *Nature* **331**, 237-249 (1988).
13. Matz, S. M. *et al. Nature* **331**, 416-418 (1988).

Resolution of the circumstellar disk of β Pictoris at 10 and 20 μm

C. M. Telesco*, E. E. Becklin†, R. D. Wolstencroft‡ & R. Decher*

*Space Science Laboratory, NASA Marshall Space Flight Center, ES63, Huntsville, Alabama 35812, USA

†Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA

‡Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK

The A5 dwarf β Pic is a star that may be surrounded by a proto-planetary disk. Among the stars discovered with the Infrared Astronomical Satellite (IRAS) to have far-infrared excesses from circumstellar dust^{1,2}, β Pic has a disk that is apparently viewed nearly edge-on, which increases its surface brightness and spatial contrast. Using coronagraphs to image the disk at optical wavelengths, it has been found to extend to at least 1,100 AU from the star³⁻⁶ and is viewed as starlight scattered from dust. The colour of the disk light is the same as, or redder than, the direct starlight, which implies that the scattering grains are larger than $\sim 1 \mu\text{m}$ (refs 3, 4). Likewise, some models of radiative transfer in a dusty disk which are constrained by the IRAS scans and fluxes² have not required submicron-sized grains and indicate a possible dust-free zone near the star^{5,7}. Constraints on these models have been limited partly by the use at optical wavelengths of an occulting disk to block the direct starlight and by the low spatial resolution of the 12–100 μm IRAS observations, which greatly limit information about the central region within $\sim 10''$, or 150 AU, of the star. Here we present observations at 10 and 20 μm with $\sim 5''$ resolution of the central $\sim 13''$ -diameter region of β Pic; these resolve the inner disk and provide important constraints for the models and a basis for removing some of their ambiguities.

The observations of β Pic were obtained on 19, 21 and 22 August 1987 at the NASA Infrared Telescope Facility on Mauna Kea using the 20-pixel bolometer array developed at the NASA

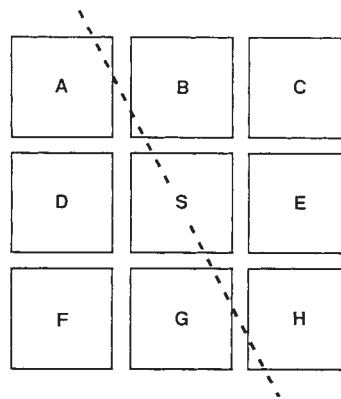


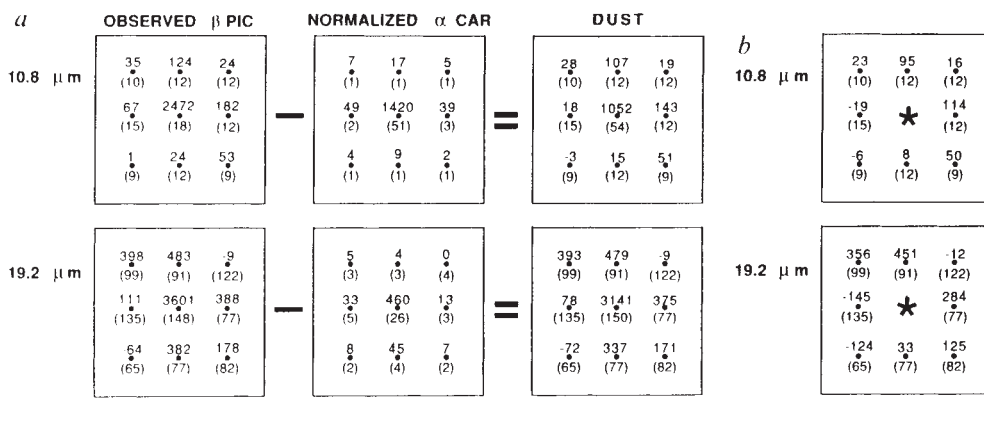
Fig. 1 The relative distribution and sizes of the nine array pixels within which β Pic was observed. All observations were made with β Pic or the standard star, α Car, centred in pixel S. The dashed line is the approximate orientation of the edge-on disk, determined from visual observations at distances $\geq 5''$ from the star. North is up and east is to the left.

Marshall Space Flight Center. The array properties and observing parameters are summarized in a footnote to Table 1. The nine array pixels to which we refer throughout this paper are indicated in Fig. 1. The star α Car, which is only $\sim 5^\circ$ from β Pic, was used as the primary photometric standard. The observations were made during early morning daylight hours and, because of the extreme southerly declination (-51°) of β Pic, at airmasses ≥ 3.1 . The observational procedure on both β Pic and α Car consisted of centering the star in a pixel near the middle of the array by maximizing the 2.2 μm signal from that pixel, followed immediately by integrations through the 10.8 μm or 19.2 μm filter. Telescope tracking, which was monitored before and after each integration, was excellent. The total integration times were 52 min at 10.8 μm and 154 min at 19.2 μm . After correction for the relative response of each pixel, flux densities for each integration set of 24 s were computed relative to the mean flux densities for eight pixels located outside the region shown in Figs 1 and 2; because of the large low-frequency background fluctuations arising from the high air masses, this technique resulted in a substantial reduction in the final computed noise.

The measured flux densities for the central nine pixels are shown in Fig. 2a. The total detected flux densities across the $13'' \times 13''$ region are 2.73 ± 0.04 Jy at 10.8 μm and 4.37 ± 0.27 Jy at 19.2 μm ; the uncertainties do not include the $\pm 10\%$ in absolute calibration. The 10.8 μm value agrees well with the flux measured by IRAS (ref. 2) at 12 μm which must therefore originate within ~ 100 AU of the star. Our flux density at 19.2 μm is consistent with a reasonable interpolation between the IRAS flux densities at 12 and 25 μm . The 10.8 and 19.2 μm flux densities detected in the central pixel are in excellent agreement with those detected in a $4''$ -diameter aperture by D. E. Backman, F. C. Gillett and F. C. Witteborn (personal communication). In Fig. 2a we have scaled the α Car fluxes so that the peak values at 10.8 and 19.2 μm equal the β Pic photospheric flux densities of 1.42 ± 0.05 Jy at 10.8 μm and 0.46 ± 0.02 Jy at 19.2 μm , determined from the photospheric values of D.E.B., F.C.G. and F.C.W. (personal communication). The α Car flux density matrix represents the response of the array to a point source centred in pixel S. The difference (Fig. 2a) between the β Pic and normalized α Car flux matrices is the distribution of β Pic flux densities that is attributable to emission from warm dust; these values would be observed in the absence of the β Pic photosphere.

Figure 2a indicates firstly that the dust emission is strongly peaked in pixel S, and therefore that the dust is primarily

Fig. 2 *a*, Flux densities at 10.8 μm and 19.2 μm for β Pic and α Car. Measurement uncertainties (one standard deviation of the measured mean) are given in parentheses and do not include the $\pm 10\%$ in calibration. Flux densities for α Car, assumed to be a point source, have been scaled so that the peak flux equals the value of the β Pic photospheric flux. The difference between the β Pic and α Car flux densities at each position is attributed to circumstellar dust; *b*, emission from dust in the outer pixels after subtraction of the flux resulting from the bright concentration of dust at the location of the central pixel.



confined to the region within $\sim 3''$ of the star, and secondly that significant emission from dust is detected outside the central pixel. To estimate the emission from dust in the region outside the central pixel, we must correct the flux in each of the outer pixels for that resulting from the detection of the bright central dust concentration. We do this by normalizing the peak α Car fluxes to the peak circumstellar dust values in Fig. 2a. The difference between the circumstellar dust and α Car emission matrices is the distribution of emission from dust in the outer eight pixels, as shown in Fig. 2b. Figure 2 indicates that the pixels with signal-to-noise ratios ≥ 3 at 10.8 μm and/or at 19.2 μm are located NE and SW of the central pixel. This distribution is consistent with its having an orientation equal to that, at PA $\sim 29^\circ$, of the large-scale edge-on disk, as determined by several groups^{4,5} (see Fig. 1). The positions where we have detected mid-infrared radiation in β Pic are distributed roughly along the edge-on circumstellar disk, and we conclude that we have resolved the disk at mid-infrared wavelengths. To quantify this conclusion somewhat, we define the asymmetry factor $\varepsilon \equiv (A + H) - (C + F)$, where the letters refer to the flux densities of the corresponding pixels designated in Fig. 1. For a completely symmetric distribution, $A = C = F = H$ and $\varepsilon = 0$, whereas $\varepsilon > 0$ for an edge-on disk oriented at position angle $= 45^\circ$. Using the dust fluxes in Fig. 2a, we obtain $\varepsilon = 63 \pm 20$ mJy at 10.8 μm and 645 ± 189 mJy at 19.2 μm . These values deviate significantly from zero, and, considering the non-optimum relative orientation of the disk and the line connecting pixels A and H, they imply that we have resolved the disk. We also note a clear indication of a NW-SE asymmetry of the infrared distributions with respect to the visually determined disk ridge line. However, this asymmetry is probably an artefact resulting from slight positioning errors of β Pic or α Car in the central pixel. This possibility is strengthened by there being a similar asymmetry (although with higher fluxes to the SE rather than to the NW) in the 19.2 μm α Car matrix in Fig. 2a. In the analysis below we do not use the fluxes in pixels B and E.

Dust temperatures can be estimated from the observed ratios of the 10.8 μm and 19.2 μm flux densities. If the dust near β Pic contains silicates, the silicate emission bands at 10 μm and 20 μm may contribute to our detected fluxes. Such contributions can lead to the derivation of erroneous values of the temperatures. Studies of cometary grains⁸⁻¹¹ suggest that strong silicate features are not present for dust temperatures as low as those considered below, but our conclusions remain preliminary until spectroscopic observations of β Pic become available which directly address this issue.

It is immediately apparent from the dust temperatures in β Pic that the mid-infrared-emitting grains are not blackbodies, because a blackbody located at pixels A and H (see below) would have a temperature of ~ 50 K, whereas the observed

colour temperature (for wavelength-independent emissivity) is ~ 163 K. The assumption (ref. 2; P. Artymowicz, C. Burrows and F. Paresce; D.E.B., F.C.G. and F.C.W., personal communication) that the emissivity Q_ν of the mid-infrared-emitting dust grains varies as ν or ν^2 implies values of the dust temperature shown in Table 1. We present temperatures for the central pixel S and for the pixels A and H located roughly along the disk. We find marginally statistically significant ($\sim 2.3\sigma$) evidence that the dust in pixels A and H is cooler than the dust in pixel S. The temperatures and mid-infrared flux densities also indicate that the emission is optically thin; for example, $\tau \approx 3 \times 10^{-4}$ for pixel S.

If we consider the equation of radiative equilibrium for a dust grain heated by β Pic (luminosity $6.5 L_\odot$), the temperatures of the dust in pixels A and H imply Planck-averaged emissivities $\langle Q_\nu \rangle = 0.02$; we have assumed that the emissivity $Q_\nu = 1$ in the spectral region of the absorbed radiation and, in view of the steep brightness distribution, that the projected distances of A and H from the star are 81 AU, the distance to the point midway between the pixel centre and the pixel corner nearest to the star. Draine and Lee¹² have calculated Planck-averaged emissivities for astronomical silicates and graphite. Interpolating within their Fig. 11, we find that spherical dust grains at temperatures of 120–140 K have emissivities equal to our derived value if they have radii $a = 0.08$ – 0.16 μm for silicates, and 0.23 – 0.34 μm for graphite. We conclude that sub-micron-sized grains emit much of the excess 10 and 20 μm radiation from β Pic. However,

Table 1 Dust flux densities and dust temperatures

Pixels*	F_ν (10.8 μm) (mJy)	F_ν (19.2 μm) (mJy)	T_d (for $Q \propto \nu$) (K)	T_d ($Q_\nu \propto \nu^2$) (K)
S†	$1,026 \pm 72$	$2,945 \pm 386$	$175(+9, -7)$	$148(+7, -5)$
A + H	73 ± 13	481 ± 129	$139(+13, -9)$	$122(+10, -7)$

The observing parameters were as follows: the NASA-MSFC array was configured in a 5×4 (RA $\times$ Dec) arrangement; pixel size (FWHM) = $4.2'' \times 4.2''$; pixel centre-to-centre separation = $4.5''$; Chopper throw = $20''$ in RA; passband centres and widths (FWHM) were 10.8 μm (5.3 μm) and 19.2 μm (5.2 μm); the photometric standard was α Car for which F_ν (10.8 μm) = 126 Jy and F_ν (19.2 μm) = 40 Jy. Uncertainties for flux densities and temperatures given in the table do not include effects of $\pm 10\%$ uncertainty in absolute calibration. All flux densities presented in this paper have been corrected for colour effects due to finite spectral bandwidths.

* Pixels are as designated in Fig. 1.

† The flux density at each wavelength for the central pixel is the difference between the total flux density from dust in the central $13'' \times 13''$ and that (Fig. 2b) from the outer eight pixels (281 ± 22 mJy at 10.8 μm and 968 ± 272 mJy at 19.2 μm); therefore, these fluxes include a slight correction for detection by pixel S of flux originating outside pixel S.

because the dust composition is unknown, the presence of graphite-like grains as large as $2a \approx 1 \mu\text{m}$ remains consistent with our observations and analysis, although much larger grains must be unimportant contributors to the 10 and 20 μm emission. Others have independently come to similar conclusions (D.E.B., F.C.G. and F.C.W., personal communication). Because visible light scattered by grains with a $\leq 1 \mu\text{m}$ is made bluer (ref. 3; P.A., C.B. and F.P., personal communication), our considerations therefore suggest that the mid-infrared-emitting dust is not the same population responsible for the neutral-colour⁴ or red³ scattering observed at visible wavelengths. The total mass of dust emitting at 10 and 20 μm across the central 13"-diameter region is $\sim 10^{22}$ g, assuming $a = 0.2 \mu\text{m}$, density $\rho = 2 \text{ g cm}^{-3}$, $\langle Q_e \rangle = 0.02$, and a dust temperature of 160 K, roughly appropriate to the total detected dust fluxes.

The Poynting-Robertson drag¹³ on sub-micron-sized grains can greatly limit their lifetimes. Grains with $a = 0.2 \mu\text{m}$ and located ~ 80 AU from β Pic should spiral into the star in $\sim 3 \times 10^5$ yr, which implies that the dust must be replenished^{2,5}. Ferlet *et al.*¹⁴ provide evidence for infall of matter onto β Pic, consistent with the picture of a dynamically evolving disk. A rough estimate of the replenishment rate can be derived by assuming that the $\sim 10^{22}$ g of 0.2- μm grains are replenished every $\sim 10^5$ yr: the rate is $\sim 10^9 \text{ g s}^{-1}$. Comets are a possible dust reservoir in β Pic^{14,15}. Dust mass-loss rates from comets in our solar system are typically less than 10^5 – 10^6 g s^{-1} for a comet at a heliocentric distance greater than ~ 1 AU (refs 10 and 16). Therefore, $> 10^3$ – 10^4 such comets would be needed to replenish the sub-micron-sized grains in β Pic. We point out, however, that stellar radiation pressure on grains in the 0.1–1 μm size range could substantially decrease their loss rate from the β Pic disk because of Poynting-Robertson drag (D.E.B., F.C.G. and F.C.W., personal communication; refs 13 and 17).

Finally, we note that the value of the Planck-averaged emissivity derived from the emission in pixels A and H, if applicable to the dust in pixel S, implies that the dust there cannot be much closer than ~ 50 AU to the star without the grain temperatures exceeding those inferred by us (Table 1). Grains at that distance, and displaced from the star perpendicular to our line of sight would lie near the outer edge of the central pixel. Therefore, our observations strengthen the conclusion⁵ that β Pic is surrounded by a region, possibly extending out to several tens of AU, that is relatively devoid of dust grains.

We acknowledge useful comments and/or data in advance of publication from Dana Backman, Humberto Campins, Martin Harwit, Mike Jura and Mike Werner, and assistance with the observations from Charles Sisk and with the data reduction from Carl Benson. We thank Dave Morrison and the IRTF staff for support, and Bill Golisch, the IRTF telescope operator, whose special efforts were essential to the successful completion of this program. The authors were guest investigators at the NASA Infrared Telescope Facility which is operated by the University of Hawaii under contract to the National Aeronautics and Space Administration.

Received 28 April; accepted 15 July 1988.

1. Aumann, H. H. *Pubs astr. Soc. Pacif.* **97**, 885–891 (1985).
2. Gillett, F. C. in *Light on Dark Matter* (ed. Israel, F. P.) 61–69 (Reidel, Dordrecht, 1985).
3. Gradije, J., Hayashi, J., Zuckerman, B., Epps, H. & Howell, R. in *Proc. 18th Lunar and Planetary Conf.* **1** 351 (1987).
4. Paresce, F. & Burrows, C. *Astrophys. J. Lett.* **319**, L23–25 (1987).
5. Smith, B. A. & Terrielle, R. J. *Science* **226**, 1421–1424 (1984).
6. Smith, B. A. & Terrielle, R. J. *Bull. Am. Astr. Soc.* **19**, 829 (1987).
7. Diner, D. J. & Appleby, J. F. *Nature* **322**, 436–438 (1986).
8. Rieke, G. H. & Lee, T. A. *Nature* **248**, 737–740 (1974).
9. Tokunaga, A. T., Golisch, W. F., Griep, D. M., Kaminski, C. & Hanner, M. S. *Astr. J.* **92**, 1183–1190 (1986).
10. Ney, E. P. in *Comets* (ed. Wilkening, L. L.) (Univ. Arizona Press, Tucson) 323–340 (1982).
11. Hanner, M. S. *Adv. Space Res.* **5**, 325–334 (1986).
12. Draine, B. T. & Lee, H. M. *Astrophys. J.* **285**, 89–108 (1987).
13. Burns, J., Lamy, P. L. & Soter, S. *Icarus* **40**, 1–48 (1979).
14. Ferlet, R., Hobbs, L. M. & Vidal-Madjar, A. *Astr. Astrophys.* **185**, 267–270 (1987).
15. Harper, D. A., Loewenstein, R. F. & Davidson, J. A. *Astrophys. J.* **285**, 808–812 (1984).
16. Hanner, M. S. *Adv. Space Res.* **4**, 189–196 (1984).
17. Harwit, M. J. *geophys. Res.* **68**, 2171–2180 (1983).

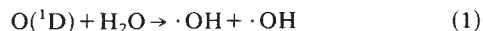
Measurements of the diurnal OH cycle by a ¹⁴C-tracer method

Colin C. Felton*, J. C. Sheppard* & M. J. Campbell†

* Department of Chemical Engineering and † Laboratory for Atmospheric Research, Washington State University, Pullman, Washington 99164, USA

The hydroxyl free radical, OH, is generally recognized as the primary oxidant for the removal of pollutants from the Earth's atmosphere. Measurements of OH concentrations are needed to test models of atmospheric photochemistry. We have used the ¹⁴C-tracer technique to measure five ambient diurnal OH concentration cycles in relatively pure and in polluted air. The early-to mid-October maximum midday OH concentrations for pure and polluted air were found to be, respectively, 2.4×10^6 and $9.5 \times 10^6 \text{ radicals cm}^{-3}$ (10^{-13} – 4×10^{-13} relative to air). Night-time OH concentrations of less than $2 \times 10^5 \text{ radicals cm}^{-3}$ were measured. Estimates of OH concentrations from photochemical models and trace gas lifetimes are consistent with our observations.

The OH radical is accepted as the primary atmospheric oxidizing agent¹, and model calculations indicate that peak OH concentrations range between 10^6 and $10^7 \text{ radicals cm}^{-3}$, five or six orders of magnitude lower than the more abundant but less reactive oxidant, O₃. Sources of atmospheric OH are believed to be



with O(¹D) formed by ultraviolet photolysis of ozone. Other sources include photolysis of H₂O₂, HNO₂, CH₂O and



Because of the daily ultraviolet light intensity variation, peak OH concentrations are expected near noon. Photochemical models of clean air chemistry^{2,3} indicate that OH is consumed by reaction with NO₂, H₂, CH₄ and other species, but most significantly with CO.

Much effort has been expended to measure OH concentrations by three methods: laser-induced fluorescence (LIF), long-path light absorption (LPA) and ¹⁴C tracers. Investigators using these techniques have encountered difficulties arising from the exceedingly low concentrations, the high reactivity and the short lifetime (~ 0.1 s) of the OH radical.

The trace-gas lifetime method, first applied by Singh *et al.*⁴ and followed by many others^{5,6}, has yielded global annual average OH concentrations of about $6 \times 10^5 \text{ radicals cm}^{-3}$ for the entire atmosphere. The primary disadvantage of this method is that it cannot determine local OH concentrations at a specified site and time.

The LIF technique, first attempted by Baardson and Terhune⁷, refined by Wang and coworkers^{8,9} and Davis *et al.*^{10,11} and recently further improved by the use of low-pressure techniques^{12–14} uses the highly selective excitation of OH at 282 nm and subsequent fluorescence at 309 nm. This promising method is limited by low photon counting efficiencies, potential saturation effects¹⁵, base line shifts¹³ and production of OH by laser-induced photolysis of O₃ (refs 15–17). Currently the LIF method requires signal integration times of more than an hour, which obscures short-term fluctuations of the OH concentration.

The LPA technique, developed by Perner and coworkers^{2,18,19}, is a simple, absolute method using absorption by OH at 282 nm, along with very long path lengths (6–10 km), permitting measurement of daily OH concentration cycles. The technique requires long signal integration times, and both absorbing impurities such as SO₂, and atmospheric turbulence, are interferences which cause a high detection limit of about $10^6 \text{ OH radicals cm}^{-3}$. Again, long integration times obscure short-term fluctuations of OH concentrations.

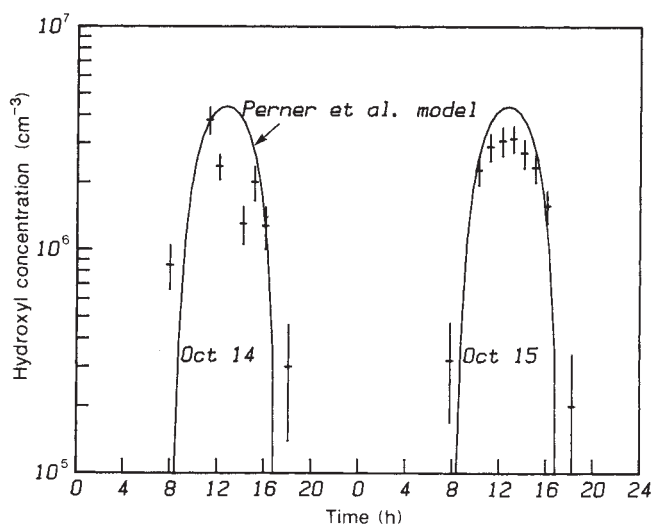


Fig. 1 Clean-air hydroxyl measurements at the Colfax clean-air site in Washington. The solid line shows the concentrations expected from the model of Perner *et al.* (2) with the data of Table 1 and 200 p.p.t. NO_x . Error bars are 1 standard deviation.

The basic assumptions of the ^{14}C method are that CO is oxidized almost exclusively by OH (refs 20–22), that the CO–OH rate constant is well characterized²³ and that the air inside the flow reactor is in the same photostationary state as the air being sampled. Interferences due to radiolysis²⁴ and surface catalysis²⁵ can be shown to be small. Further, the $^{14}\text{CO}_2$ must also be separated effectively from radioactive contaminants such as HTO, $^{14}\text{CH}_4$, CH_3T , ^{85}Kr and ^{222}Rn (ref. 26). These are very stringent requirements and have been very difficult to satisfy. The method employs ^{14}CO as a tracer in a cylindrical continuous-flow quartz reactor, which is transparent to ultraviolet light and does not alter the ambient ultraviolet flux²⁷. The reactor is designed with flow characteristics which minimize perturbations of the photochemistry of the air being studied²⁸. The ^{14}CO is injected into air which is drawn through the reactor and the resulting $^{14}\text{CO}_2$ is collected cryogenically and measured. At any point in the reactor

$$[^{14}\text{CO}_2] = \int_0^t k [^{14}\text{CO}] [\text{OH}] dt. \quad (2)$$

Assuming that the ^{14}CO is negligibly consumed in the reactor, that the $^{14}\text{CO}_2$ and ^{14}CO are dispersed at equal rates in the tube and that OH is not depleted significantly in the reactor due to ^{14}CO , then:

$$[\text{OH}] = \frac{[^{14}\text{CO}_2]}{[^{14}\text{CO}]kt} \quad (3)$$

where $k = \text{CO–OH}$ reaction rate constant ($2.3 \times 10^{-13} \text{ cm}^3 \text{ molecules}^{-1} \text{ s}^{-1}$), $t = \text{mean } ^{14}\text{CO}$ residence time in reactor (reaction time), and concentrations refer to those at the end of the reactor.

For the desired OH sensitivity a reaction time of 12 s was chosen (flow velocity = 6.7 cm s^{-1}) and the $^{14}\text{CO}_2$ was collected for 100 s. To measure the concentration of $^{14}\text{CO}_2$ at the end of the reactor, air from the central portion of the tube is collected to eliminate the effects of wall reactions and cooled with liquid oxygen to condense the CO_2 . Another air sample is taken to measure the ^{14}CO concentration. The $^{14}\text{CO}_2$ is subsequently separated from the unreacted ^{14}CO using isotope dilution and from ^{222}Rn by the use of a column packed with porous polymer beads. In the past ^{222}Rn produced highly variable backgrounds in the instrument^{22,29}. Counting of the $^{14}\text{CO}_2$ and ^{14}CO samples was done in the Washington State University Radiocarbon Dating Laboratory³⁰.

The basis of the method, ^{14}CO oxidation in a continuous reactor, is the same as that used by Campbell *et al.*^{22,29} in OH

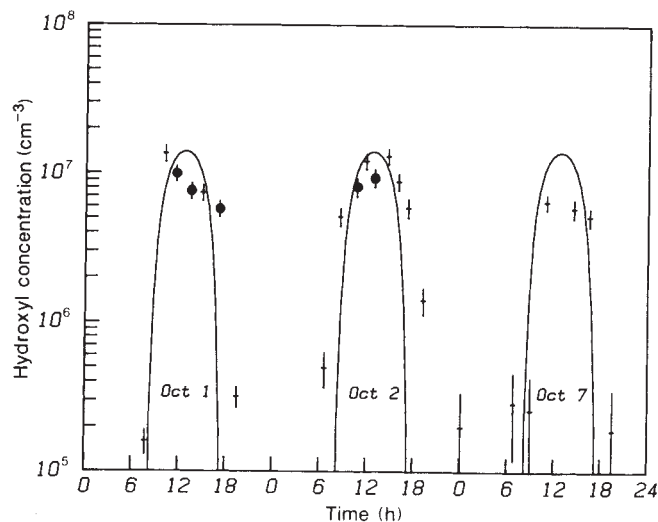


Fig. 2 Hydroxyl measurements in air contaminated by a power plant plume. The solid lines are model calculations, scaled in magnitude to match the data, indicating the form of the diurnal cycle. Filled circles denote results from a reactor shielded from photolytic radiation. Error bars are 1 standard deviation. Note the low-light OH concentrations of $\sim 2 \times 10^5 \text{ cm}^{-3}$ reflecting typical instrumental backgrounds.

intercomparison tests at Wallops island, Virginia. The purification system and instrument operation are, however, quite different. The Pullman experiments were performed on 1, 2 and 7 October in fumigating plumes of a nearby coal-fired power plant and represent tests of recent improvements to the apparatus. The results of 14 and 15 October represent more controlled tests in a rural area near Colfax, Washington.

Auxiliary data in Table 1 include ambient CO , CH_4 , H_2O , O_3 , NO_x ($\text{NO} + \text{NO}_2$) concentrations and theoretical $\text{O}(^1\text{D})$ photoproduction rates³¹. The NO_x concentrations were usually below the detection limit of the instrument.

Comparison of results shows, in the results for 1, 2 and 7 October, the effect of an elevated albedo ($\sim 50\%$) on the roof of Dana Hall, Pullman, and of increased NO_x and H_2O concentrations. Noontime OH levels on 1, 2 and 7 October were about 4 times the rural concentrations on 14 and 15 October, possibly because of the transient effects of $\text{NO}_x + \text{HO}_2$ reaction in the power plant plume. The effect of cloud cover can be seen in comparing the data for 14 and 15 October.

The standard deviations in Figs 1 and 2 include all significant errors associated with these measurements. Collection efficiencies for the $^{14}\text{CO}_2$ have a $\pm 10\%$ uncertainty, but the largest errors were in the $^{14}\text{CO}_2$ activity determinations, which are proportional to the square root of the counting time.

An error not considered was the effect of residual ^{222}Rn in the $^{14}\text{CO}_2$ sample. Depending on the ambient ^{222}Rn concentration, the background activity of the collected CO_2 was between 0.02 and 0.1 Bq and was confirmed by experiments in which

Table 1 Average auxiliary data

Date	CO (p.p.b.)	CH ₄ (p.p.m.)	NO _x * (p.p.b.)	H ₂ O (%)	¹⁴ CO (p.p.b.)	Peak JO(¹ D)† (s ⁻¹ × 10 ⁵)	O ₃ (p.p.b.)
1-10-87		2.0	<5–100	0.56	69	2.3	35
2-10-87		1.7	<5–40	0.53	27	2.2	35
7-10-87	428	1.8	<5–175	0.52	23	2.1	50
14-10-87	190	1.8	<5	0.42	24	1.2	38
15-20-87	153	1.8	<5	0.42	28	1.2	35

p.p.b., parts per 10⁹.

* NO_x was below instrument detection limit except for spikes of $\sim 30 \text{ s}$ duration.

† Photolysis rates scaled from values obtained using the correlations of Demerjian *et al.* for Colfax.

^{14}CO was not injected into the reactor. Correction for this relatively small effect can be made by subtracting the lowest OH concentrations measured that day from all the higher concentrations, because models consistently predict night-time OH concentrations less than 10^5 radicals cm^{-3} . These corrections, usually equivalent to less than 2.5×10^5 radicals cm^{-3} , have not been made in the present data. The precision of the results can also be estimated by comparing the data for 1, 2 and 7 October with those for 14 and 15 October; very similar results were obtained on days with comparable meteorological and chemical conditions.

These results are significant because the low $^{14}\text{CO}_2$ levels in the early morning and evening suggests that the elevated midday $^{14}\text{CO}_2$ levels are caused by the photochemical reaction between CO and OH. Experiments in which the reactor was shielded from ultraviolet radiation (Fig. 1) yielded OH concentrations which were slightly smaller than adjacent unshielded values, suggesting that the primary OH source is the $\text{HO}_2\text{--NO}$ reaction. Weinstock *et al.*³² show that the OH concentrations do not decay completely within the reaction time upon ultraviolet flux interruption.

A more detailed presentation of the method, results and model comparisons will be made elsewhere. Future improvements for this method include increased separation of ^{222}Rn from the $^{14}\text{CO}_2$ and reduction of the sample integration time to 25 s. With complete radon removal, a detection limit of 10^4 radicals cm^{-3} is feasible, enabling night-time detection of OH. In contrast to laser techniques, limitations of the radiochemical technique are determined by operations performed in the laboratory. Improvements in separation efficiency can be made at low cost and will result in a proportionately lower detection limit until counting becomes limiting at about 10^4 radicals cm^{-3} . Laser methods, however, are limited primarily by uncontrollable atmospheric fluctuations. Also, the integration time of 100 s is much less than that required for laser methods^{14,15,17,18}, which average signals for several hours in the course of a single OH measurement. Just as local measurements are preferable to global measurements in elucidating basic atmospheric chemistry, there is great value in a measurement time approaching the lifetime of the species of interest.

This work was partially supported by the Washington State University College of Engineering, the NSF and NASA grants. We also acknowledge the help of Harbi Elshafei, J. Carl Farmer and Yvonne Welter.

Received 18 April; accepted 18 July 1988.

- Levy, H. *Science* **173**, 141–143 (1971).
- Perner, D. *et al. J. atmos. Chem.* **5**, 185–216 (1987).
- Logan, J. A., Prather, M. J., Wofsy, S. C. & McElroy, M. B. *J. geophys. Res.* **86**, 7210–7254 (1981).
- Singh, H. B. *Geophys. Res. Lett.* **4**, 453–456 (1977).
- Neely, W. B. & Plonka, J. H. *Envir. Sci. Technol.* **12**, 317–321 (1978).
- Fraser, P. J., Hyson, P., Rasmussen, R. A., Crawford, A. J. & Khalil, M. A. K. *J. atmos. Chem.* **4**, 3–42 (1986).
- Baardson, E. L. & Terhune, R. W. *Appl. Phys. Lett.* **21**, 209–211 (1972).
- Wang, C. C. & Davis, L. I. *Phys. Rev. Lett.* **32**, 349–352 (1974).
- Davis, L. I. *et al. J. geophys. Res.* **92**, 2020–2024 (1987).
- Davis, D. D., Heaps, W. & McGee, T. *Geophys. Res. Lett.* **3**, 331–333 (1976).
- Davis, D. D., Heaps, W., Philen, D. & McGee, T. *Atmos. Envir.* **13**, 1197–1203 (1979).
- Hard, T. M., O'Brien, R. J., Cornelius, Y. C. & Mehrabzadeh, A. A. *Envir. Sci. Technol.* **18**, 768–777 (1984).
- Hard, T. M., Chan, C. Y., Mehrabzadeh, A. A., Pan, W. H. & O'Brien, R. J. *Nature* **322**, 617–620 (1986).
- Shirinzadeh, B., Wang, C. C. & Deng, D. Q. *Geophys. Res. Lett.* **14**, 123–126 (1987).
- Ortigue, G., Gericke, K.-H. & Comes, F. J. *Z. Naturf.* **36a**, 177–183 (1981).
- Wang, C. C., Davis, L. I., Selzer, P. M. & Munoz, R. J. *geophys. Res.* **86**, 1181–1186 (1981).
- Davis, D. D., Rodgers, M. O., Fischer, S. D. & Heaps, W. S. *Geophys. Res. Lett.* **8**, 73–76 (1981).
- Perner, D. *et al. Geophys. Res. Lett.* **8**, 466–468 (1976).
- Platt, U., Rateike, M., Junkermann, W., Hofzumahaus, A. & Ehhalt, D. H. *Free Radical Res. Commun.* **3**, 165–172 (1987).
- Campbell, M. J., Sheppard, J. C. & Au, B. F. *Geophys. Res. Lett.* **6**, 175–178 (1979).
- Sheppard, J. C., Hardy, E. J. & Hopper, F. *Antarct. J. U.S.* **17**, 206–207 (1982).
- Campbell, M. J. *et al. J. atmos. Chem.* **4**, 413–427 (1987).
- Hynes, A. J., Wine, P. H. & Ravishankara, A. R. *J. geophys. Res.* **91**, 11815–11820 (1986).
- Hardy, R. J., Sheppard, J. C. & Campbell, M. J. *Int. J. appl. Radiat. Isotopes* **35**, 1071–1072 (1984).
- Farmer, J. C., Fitzner, C. A., Campbell, M. J., Henry, M. N. & Sheppard, J. C. *Int. J. appl. Radiat. Isotopes* **36**, 915–918 (1985).

- Felton, C. C. thesis, Washington State Univ. (1988).
- Zafonte, L., Rieger, P. L. & Holmes, J. R. *Envir. Sci. Technol.* **11**, 483–487 (1977).
- Meyn, L. A. thesis, Washington State Univ. (1982).
- Hoell, J. M. *et al. J. geophys. Res.* **89**, 11819–11825 (1984).
- Sheppard, J. C. College of Engineering Bulletin 338, Washington State Univ. (1975).
- Demerjian, K. L., Schere, K. L. & Peterson, J. T. *Adv. Envir. Sci. Technol.* **10**, 369–459 (1980).
- Weinstock, B., Niki, H. & Chang, T. Y. *Adv. Envir. Sci. Technol.* **10**, 221–258 (1980).

Surface electronic properties probed with tunnelling microscopy and chemical doping

Xian-Liang Wu, Peng Zhou & Charles M. Lieber*

Department of Chemistry, Columbia University,
New York 10027, USA

Scanning tunnelling microscopy^{1–3} (STM) can provide atomic-resolution images of surfaces in vacuum^{4,5}, air⁶ and liquids^{7,8}. One of the most appealing aspects of such images is that they appear to reflect surface structure directly; these tunnelling images, however, contain contributions from both the structural and the electronic properties of a surface^{1–5}. Although an understanding of these properties is essential to an understanding of the fundamental nature and reactivity of surfaces, few methods⁵ are available to separate them, especially in air and in liquids. Here we report a new approach to this problem that combines chemical modifications with tunnelling microscopy. Samples of the layered material tantalum disulphide (TaS_2) have been substitutionally doped with titanium to prepare materials of the general form $\text{Ti}_x\text{Ta}_{1-x}\text{S}_2$. STM images of native TaS_2 are dominated by a charge density wave state^{9–11}. Using titanium doping, we have been able to perturb this unusual electronic feature systematically so that the surface structure can be imaged clearly. Such studies of chemically modified materials (prepared, for example by doping or intercalation) should lead to a better understanding of the features contained in STM images.

Tantalum disulphide is a member of the transition metal dichalcogenides, a group of layered materials that exhibit interesting electrical, magnetic and catalytic properties^{12,13}. The highly anisotropic structure of TaS_2 consists of covalent S–Ta–S layers that are bonded together weakly, predominantly by van der Waals interactions (Fig. 1a). The sulphur and tantalum centres are arranged in hexagonal close-packed (h.c.p.) planes with $a_0 = 3.35$ Å. Our studies concentrate on crystals in which the tantalum centres are coordinated in octahedral holes, although octahedral or trigonal prismatic coordination is possible. A unique feature of octahedrally coordinated TaS_2 is its charge density wave (CDW) state, a temperature-dependent periodic distortion of the lattice and the electron density within the S–Ta–S layers. The lattice distortions are small (<0.2 Å) but the periodic variation of the electron density is quite large (11.7 Å) (ref. 13).

Single crystals of TaS_2 , grown by chemical vapour transport using I_2 , can be cleaved along the a -axis to expose large atomically flat regions of h.c.p. sulphur atoms⁹. A typical tunnelling image of such a surface, recorded in the constant current mode using our STM, is shown in Fig. 2. A hexagonal lattice is prominent in this surface image, but the peak spacing and corrugation are 11.8 ± 0.2 and 5.7 ± 0.5 Å respectively, much larger than the values expected for the h.c.p. sulphur plane (3.35 and <1 Å respectively)¹⁴. The features observed in this STM image are due to the periodic variation in charge associated with the CDW state in TaS_2 ^{9–11}. Although there is no atomic structure visible in this image, we and others have observed such structure occasionally in other experiments (Lieber *et al.*,

* To whom correspondence should be addressed.

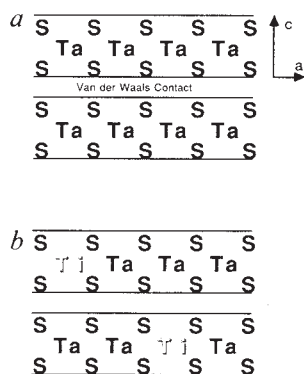


Fig. 1 Cross-sectional view of the structure of tantalum disulphide which displays two S-Ta-S layers in van der Waals contact. *b*, Tantalum disulphide substitutionally doped with titanium. The titanium-doped material is prepared by reaction of the appropriate ratio of metals and sulphur ($x\text{Ti} + (1-x)\text{Ta} + 2\text{S}$) at 950 °C.

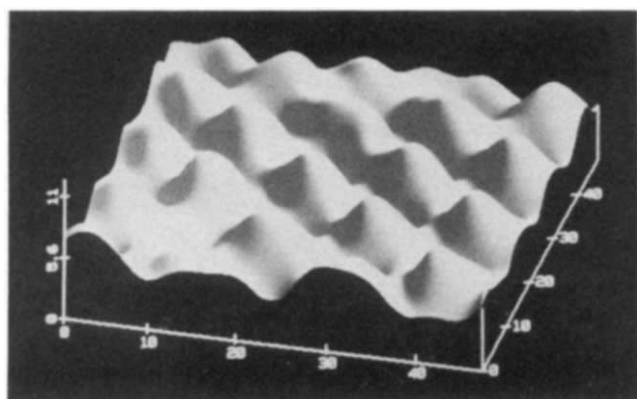


Fig. 2 STM image of TaS_2 recorded at room temperature using a modified commercial instrument (Digital). The image was recorded in the constant current mode (tunnelling current 2.1 nA) with a platinum-iridium alloy (90%–10%) tip biased at +26 mV relative to the sample. The sample surface was covered with oil (Fluorolube, Fisher Scientific), although similar images were also obtained in air. The raw digital data was low-pass-filtered before display. The axes markings are in Å.

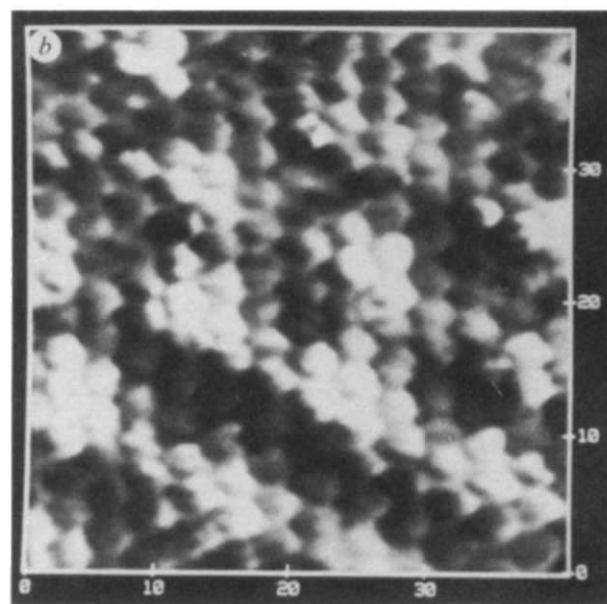
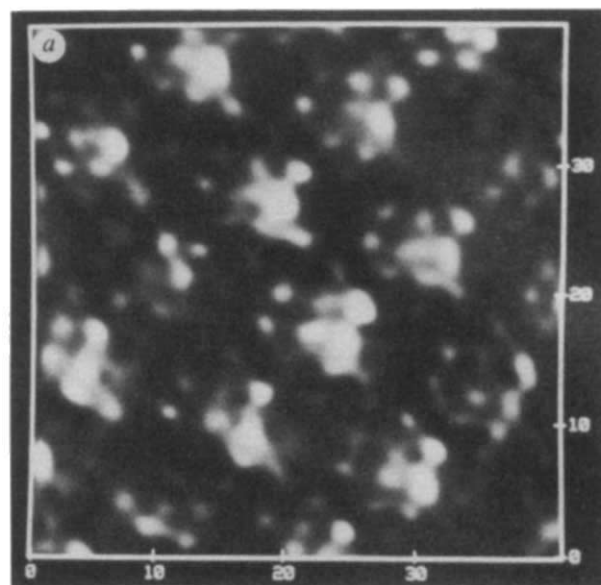


Fig. 3 Grey-scale images of $\text{Ti}_x\text{Ta}_{1-x}\text{S}_2$ recorded in the constant current mode. The sample surfaces were covered with oil to improve the imaging stability¹⁷. *a*, $x = 0.06$; tip bias = +8 mV, tunnelling current = 2.1 nA. The grey-scale in this image (black minimum to white maximum) is 4.3 Å. *b*, $x = 0.1$; tip bias voltage +9 mV, tunnelling current = 2.1 nA. The grey-scale in this image, 2.9 Å, is significantly smaller than in *a*. The x , y -axes markings are in Å.

Clarke *et al.* and Coleman *et al.*, unpublished results); nevertheless, it is difficult to observe the atomic lattice directly because of the strong CDW signal.

Knowledge of both the electronic and structural properties of a surface are necessary to understand certain characteristics, such as its reactivity. Spectroscopic tunnelling measurements have been used successfully to delineate these properties in ultrahigh vacuum⁵, but such measurements are difficult to carry out in air and liquid environments. To obtain this crucial information we have been developing methods based on well-defined chemical modifications that perturb the electronic properties of the surface without changing its structure¹¹. By imaging a series of chemically modified materials one can obtain a more complete understanding of the electronic and structural nature of a surface.

Randomly doped samples of $\text{Ti}_x\text{Ta}_{1-x}\text{S}_2$, where $x = 0.06$ and 0.1 (ref. 15), are used here (Fig. 1*b*). Although titanium doping decreases the average electron density by one electron per Ti atom, the structure of the modified material is virtually the same as the parent TaS_2 . STM images of these two samples are shown in Fig. 3. These images are typical of the many (>40) we have obtained at different locations on several different samples using different tips. It is readily apparent from Fig. 3 that even low levels of Ti-doping perturb the CDW state sufficiently that the atomic lattice can be 'seen' clearly. For the $x = 0.06$ sample, the

CDW is still prominent as large peaks (11.6 ± 0.2 Å peak-to-peak separation and 3.6 ± 0.5 Å corrugation), but examination also reveals smaller satellite peaks with separation 3.3 ± 0.1 Å, which we assign to the h.c.p. sulphur atoms. For $x = 0.1$ samples (Fig. 3*b*) the atomic lattice, with peak-to-peak separation 3.4 ± 0.1 Å, dominates the tunnelling images; the CDW is visible only as a weak periodic ~ 11.6 Å modulation of the sulphur atoms (1.4 ± 0.4 Å corrugation). The corrugations measured indicate that the CDW amplitude decreases substantially with increasing dopant concentration. Because some variation in corrugation is found for a given titanium concentration, we have analysed the corrugations from at least 15 images for each dopant concentration to determine whether this trend is significant. The images were recorded with similar instrumental parameters (bias voltage and tunnelling current). We find that the corrugations ($\pm 1\sigma$) for

$x = 0, 0.06$ and 0.10 are 5.2 ± 0.9 , 3.5 ± 0.6 and 1.2 ± 0.3 Å respectively, which demonstrates that the observed effect of Ti-doping is greater than the sample-to-sample variation for a given titanium concentration. In related work Coleman *et al.* have shown¹⁶ that the CDW amplitude decreases with increasing tunnelling current, but even at high currents no atomic structure was reported.

Although Ti-doping clearly modifies the electronic properties of TaS₂ sufficiently to enable STM to probe the surface structure, the origin of these effects is not certain. Doping with titanium causes randomness in the lattice potential and decreases the conduction-electron density by one electron per Ti atom (substituting d⁰-Ti for d¹-Ta)^{12,13}. These two effects will perturb the CDW in different ways. Disorder in the potential suppresses the temperatures at which the CDW, which is incommensurate with the lattice at high temperatures, undergoes transitions first to a quasicommensurate and then to a commensurate state. For $x = 0, 0.06$ and 0.1 the incommensurate to quasicommensurate transition temperatures are $\sim 350, 305$ and 250 K, respectively¹⁵. Hence, at our experimental temperature (295 K) the CDW is quasicommensurate with the lattice of TaS₂, close to a transition for Ti_{0.06}Ta_{0.94}S₂, and incommensurate with the lattice of Ti_{0.1}Ta_{0.9}S₂. Because the CDW amplitude increases in the quasicommensurate state¹³, we believe that the increase in observed atomic structure with increasing Ti is probably due to a decrease in the CDW amplitude associated with the incommensurate state. We are engaged in experiments designed to probe this point further.

On the other hand, a decrease in the conduction electron density will cause the Fermi surface to shrink (in the rigid-band approximation) and hence the CDW wavelength should increase^{12,13}. Diffraction studies¹⁵ suggest that for the $x = 0.1$ sample the CDW modulation should be 12.5 Å (as opposed to 11.7 Å for $x = 0$), but we observe no such change. It is not clear why the change in conduction electron density does not appear to affect the wavelength of the CDW determined from the STM images. But we note that our STM results probe the local charge modulation directly whereas diffraction studies measure the lattice distortions that are coupled to the CDW. Clearly more work is needed to understand these effects.

In conclusion, our results demonstrate how chemical modifications can be combined with tunnelling microscopy to separate the electronic and structural components of STM surface images. Such well defined chemical methods should be useful for studies of other low-dimensional materials (such as graphite and M_xWO₃) that can be modified by intercalation, doping and other methods. Such studies are currently in progress in our laboratory.

We thank F. J. DiSalvo and C. J. Chen for discussion, the referees for bringing to our attention unpublished work, and J. V. Waszczak for supplying titanium-doped samples. This work was partially supported by Presidential Young Investigator and Dreyfus New Faculty Awards (C.M.L.).

Received 30 May; accepted 20 July 1988.

- Binnig, G. & Rohrer, H. *Angew. Chem., int. Ed. Engl.* **26**, 606-614 (1987).
- Hansma, P. K. & Tersoff, J. *J. appl. Phys.* **61**, R1-R23 (1987).
- Quate, C. F. *Physics Today* **39**, 26-33 (1986).
- Becker, R. S. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4667-4670 (1987).
- Tromp, R. M., Hamers, R. J. & Demuth, J. E. *Science* **234**, 304-309 (1986).
- Park, S.-I. & Quate, C. F. *Appl. Phys. Lett.* **48**, 112-114 (1986).
- Sonnenfeld, R. & Hansma, P. K. *Science* **232**, 211-213 (1986).
- Giambattista, B. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 4671-4674 (1987).
- Coleman, R. V., Drake, B., Hansma, P. K. & Slough, G. *Phys. Rev. Lett.* **55**, 394-397 (1985).
- Slough, C. G., McNairy, W. W., Coleman, R. V., Drake, B. & Hansma, P. K. *Phys. Rev. B* **34**, 994-1005 (1986).
- Wu, X.-L. & Lieber, C. M. *J. Am. chem. Soc.* **110**, 5200 (1988).
- Wilson, J. A., DiSalvo, F. J. & Mahajan, S. *Adv. Phys.* **24**, 117-201 (1975).
- DiSalvo, F. J. in *Electron-Phonon Interactions and Phase Transitions*, (ed. Riste, T.) 107-136 (Plenum, New York, 1977).
- Slough, C. G. *et al. Phys. Rev. B* **37**, 6571-6574 (1988).
- DiSalvo, F. J., Wilson, J. A., Bagley, B. G. & Waszczak, J. V. *Phys. Rev. B* **12**, 2220-2235 (1975).
- Coleman, R. V. *et al. J. Vac. Sci. Technol. A* **6**, 338-343 (1988).
- Schneir, J. & Hansma, P. K. *Langmuir* **3**, 1025-1027 (1987).

Contaminated aquifers are a forgotten component of the global N₂O budget

Daniel Ronen*†, Mordechai Magaritz* & Ehud Almon*

* Isotope Department, The Weizmann Institute of Science, 76100 Rehovot, Israel

† Research Department, Hydrological Service, PO Box 6381, Jerusalem 91063, Israel

One of the chemical components contributing to the destruction of the ozone layer in the upper atmosphere consists of the nitrogen oxides formed from N₂O (ref. 1). Prompted by the prevailing idea that the ocean is not a major source of N₂O or a sink for N₂O, estimates have been made of global fluxes from continental ecosystems². Although most land areas are underlain by groundwater³, this medium has never been considered in global budgeting of N₂O. A large number of aquifers around the world are contaminated by nitrogen compounds, and processes of nitrification and denitrification are reported to be operative in this environment³. These processes lead to the production of N₂O (refs 4 and 5). Here we report that the concentration of N₂O in phreatic aerobic aquifers contaminated by anthropogenic activities (disposal of human or animal waste, cultivation and fertilization) are up to three orders of magnitude higher than the concentration expected as a result of equilibrium with the atmosphere.

Water samples were collected in the Netherlands and Israel (1) at the Veluwe region (52°15' N, 5°40' W), where a shallow 9 m-deep sandy aquifer under woodland is contaminated by

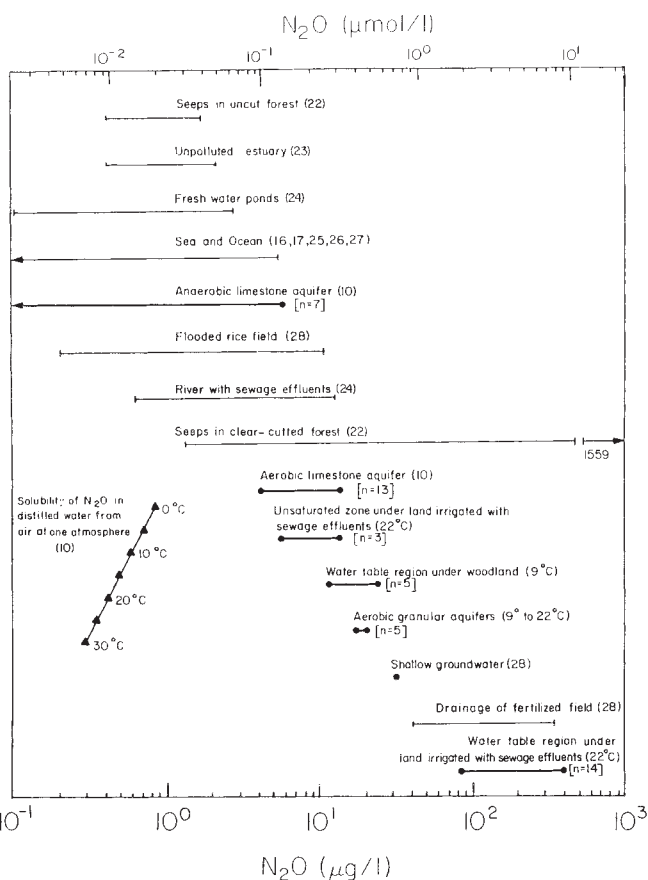
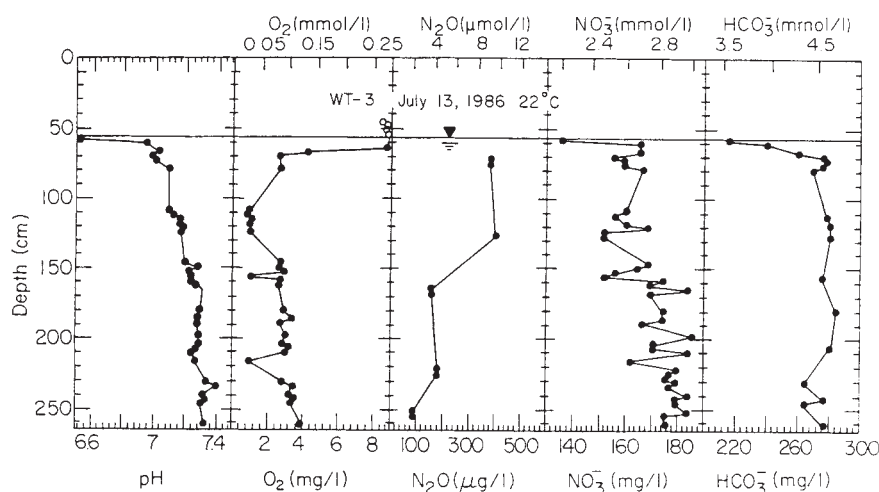


Fig. 1 Concentration range of N₂O in different aquatic systems. The theoretical solubility of N₂O at different temperatures is also given. The concentration ranges of N₂O in aquifer systems obtained during this study are denoted by filled circles. The number of samples is given by *n*. Numbers in parentheses are references.

Fig. 2 pH, O_2 , N_2O , NO_3^- and HCO_3^- concentration profiles obtained at the water-table region under land irrigated with sewage effluents (research well WT-3). The water table (denoted by the inverted triangle) is at a depth of 30 m below top soil. The vertical axis shows the depth in relation to an arbitrary reference level. The research wells are equipped with screens above groundwater. This arrangement permits the sampling of the gas composition of the unsaturated zone as it equilibrates with the distilled water in the dialysis cells. Note that the oxygen concentration found in the unsaturated zone just above the water table (open circles) reflects the expected concentration of atmospheric oxygen dissolved in water at the prevailing temperature.



acid rain, ammonia volatilization and the disposal of manure on small land enclaves⁶; (2) in the Jerusalem area (31°45' N, 35°10' W), where municipal sewage effluents infiltrate into a karstic aquifer that becomes confined and anaerobic along the groundwater flow path⁷, and (3) from two research wells (WT-2 and WT-3) in the 30 m-deep sandy Coastal Plain aquifer of Israel (32°9' N, 34°48' W) at a site where sewage effluents have been used to irrigate cultivated land for more than 20 years⁸. Studies at this location showed that dissolved organic matter deriving from the irrigation water reached the water table.

The samples were obtained from deep pumping wells and research wells. The latter were used to sample the interface between the saturated and unsaturated zone (water-table region) with a multilevel-sampler⁹. In the pumping wells, 250 ml water samples were collected in glass vessels connected by Tygon tubing to the well tap. The glass vessels were flushed with the groundwater for more than one minute before isolating the sample between two vacuum stopcocks. All wells were operated continuously for more than half an hour before sampling. Water samples in the research wells were obtained by introducing a needle through the dialysis membrane into two to three dialysis cells of the multilevel sampler and extracting the water under suction into a 25 ml glass vessel. The vessel was closed by two vacuum stopcocks after filling. All samples were immediately transported to the laboratory for N_2O extraction using a high-vacuum helium-stripping method¹⁰. The extracted gas was introduced into an 8 ml evacuated glass vial subsequently used to store and load the sample into the gas chromatograph. Nitrous oxide was analysed by electron-capture gas chromatography¹¹. The reproducibility of the gas analysis was $\pm 0.5 \mu\text{g l}^{-1}$ ($0.01 \mu\text{mol l}^{-1}$).

The concentration of N_2O in the studied aquifers is very high when compared with other unpolluted and polluted systems (Fig. 1). In the limestone aquifer the highest concentrations of N_2O are found in the aerobic section, where it may be produced from the oxidation of the organic matter influx of the sewage effluents from the city of Jerusalem¹⁰. Downstream, where the aquifer becomes confined and anaerobic, the concentration of N_2O decreases, probably as a result of its consumption during denitrification¹².

The concentration of N_2O is dramatically high at the water table region, particularly under land irrigated with sewage effluents (Fig. 1). In this region, dissolved organic carbon which trickled down through the unsaturated zone is biodegraded, consuming the dissolved oxygen found in groundwater¹³. At the prevailing minimal dissolved oxygen concentrations of about 2 mg l^{-1} (0.06 mmol l^{-1}) (ref. 13), it is reasonable to assume that N_2O is produced during nitrification. In some cases (Fig. 2) this process can be inferred from the very high concentrations of N_2O , up to $400 \mu\text{g l}^{-1}$ ($9.10 \mu\text{mol l}^{-1}$), and the concomitant

decrease in pH and bicarbonate content that could be the result of nitrogen mineralization and cell synthesis of nitrifying organisms respectively¹⁴. But most nitrate found at the water-table region is not produced *in situ* by the mineralization of nitrogen. It is the result of nitrate infiltration through the unsaturated zone, which results from intensive application of nitrogen fertilizers on the land surface¹⁵. From the total amount of dissolved organic carbon reaching the water-table region in the studied system¹³, we have calculated the N_2O/N production ratio to be 6×10^{-2} . This ratio is one order of magnitude higher than the production ratio reported for aerobic open water bodies¹⁶⁻¹⁸.

From profiles obtained under cultivated land irrigated with sewage effluents in the Coastal Plain aquifer of Israel (Fig. 3), we estimate that the N_2O flux from groundwater into the unsaturated zone ranges from 3.4 to 7.8 kg N_2O-N per hectare per year.

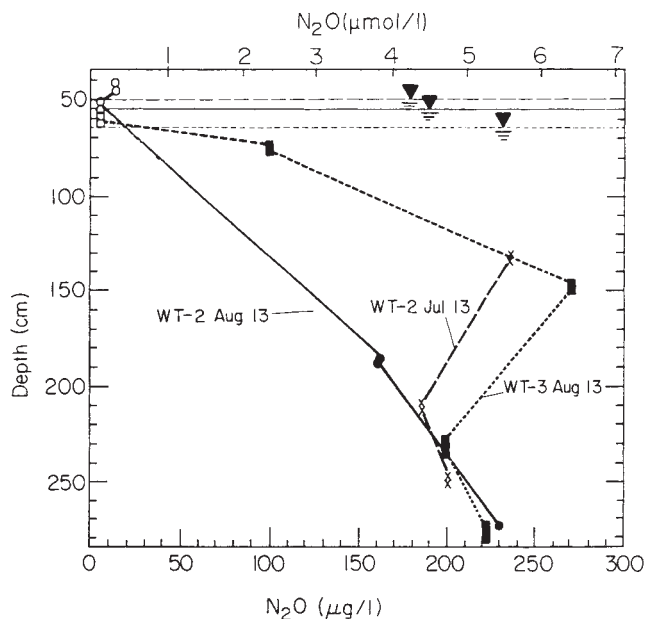


Fig. 3 Concentration profiles of nitrous oxide obtained in the water-table region of the Coastal Plain aquifer in Israel under land irrigated with sewage effluents (research well WT-2 and WT-3). Note the concentration of N_2O found in the unsaturated zone just above the water table (empty squares and circles). The water tables corresponding to the different profiles are denoted by inverted triangles. Water temperature, 22°C. The nitrous oxide flux was calculated from $\text{Flux}_{N_2O} = D_a \Psi^2 dC/dZ$, where D_a is N_2O diffusivity, $2.11 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$ at 20°C (ref. 29), Ψ the aquifer porosity, 40% (ref. 12) and dC/dZ the concentration gradient, $1.01 \text{ mg l}^{-1} \text{ cm}$ in WT-2 on 13 August and $2.33 \text{ mg l}^{-1} \text{ cm}$ in WT-3 on 13 August.

This is an extremely high emission rate when compared with other estimates from continental ecosystems². Indeed, at this site the aquifer is under a heavy organic load¹³, and the calculated N₂O fluxes are probably an upper limit. But such loads can also be expected in many aquifers on top of which organic matter from anthropogenic activities is disposed of or released from natural sources. Note also the relatively high N₂O concentrations found in the water table region under woodland that has been subjected to acid rain and ammonia volatilization from manure in the Veluwe region in the Netherlands (Fig. 1).

From the relation reported by Weiss and Price¹⁹ and using the concentration of dissolved N₂O just above the water table in the Coastal Plain aquifer of Israel (Fig. 1), it was calculated that the molar fraction of N₂O in the gas phase of the unsaturated zone is up to 12,000 parts per 10⁹. This is forty times higher than the atmospheric concentration²⁰, but we do not know if all of this will eventually reach the atmosphere after transport through the unsaturated zone. In addition, degassing of aerobic aquifers could release N₂O into the Earth's atmosphere (1) from springs and rivers that are groundwater outlets, (2) through marine environments where groundwater of coastal aquifers discharges and (3) as a result of the use of groundwater on the land surface for municipal or agricultural purposes.

Information about the concentration of N₂O in other aquifers is not available. Therefore, it is premature to calculate quantitatively the influence of aquifers on the global N₂O budget. But even a conservative estimate shows that the N₂O flux found in this study would produce 0.8×10^6 to 1.7×10^6 tons of N₂O per year globally. This value is based on the assumption that only 1% of the world groundwater sources are contaminated and that the total surface area underlain by groundwater³ is 130×10^6 km². This production rate amounts to 10–20% of the total biospheric production of N₂O (8.8×10^6 tons per year)²¹. The existing data, and the worldwide trend of increasing groundwater contamination, suggest that this ecosystem cannot be overlooked in constructing global N₂O budgets.

We thank G. J. W. Krajenbrink, Dr A. J. Amiel and S. Shasha for collaboration in field sampling, and A. Pearl for the dissolved gas analysis. Samples from pumping wells in the Netherlands were obtained by courtesy of the Water Supply Company, Gelderland. The authors also thank Drs A. Kaufmann and L. I. Gordon for discussion. This work was supported by the Israel National Council for Research and Development in the framework of the European Economic Community—Israel agreement.

Received 1 February; accepted 29 June 1988.

1. Cicerone, R. J. *Science* **237**, 35–42 (1987).
2. Banin, A., Lawless, J. G. & Whitten, R. C. *Adv. Space Res.* **4**, 207–216 (1984).
3. Freeze, R. A. & Cherry, J. A. *Groundwater* (Prentice-Hall, Englewood Cliffs, 1979).
4. Delwiche, C. C. in *Stratospheric Ozone and Man* (eds Bower, F. A. & Ward, R. B.) 65–77 (CRC, Boca Raton, Florida, 1982).
5. Burdof, J. R. & Bremner, J. M. *Soil Biol. Biochem.* **7**, 389–394 (1975).
6. Krajenbrink, G. J. W., Ronen, D., Duijvenbooden, van W., Magaritz, M. & Wever, D. J. *Hydrol.* **98**, 83–102 (1988).
7. Ronen, D., Kanfi, Y. & Magaritz, M. in *Developments in Ecology and Environmental Quality* (ed. Shoval, H. I.) 301–310 (Balaban International Science Services, Rehovot/Philadelphia, 1983).
8. Ronen, D. & Magaritz, M. *J. Hydrol.* **80**, 311–323 (1985).
9. Ronen, D., Magaritz, M. & Levy, L. *Water Res.* **20**, 311–315 (1986).
10. Almon, E. thesis, Weismann Inst. Sci. (1986).
11. Moiser, A. R. & Mack, L. *Soil Sci. Soc. Am. J.* **44**, 1121–1123 (1980).
12. Cohen, Y. *Nature* **272**, 235–237 (1978).
13. Ronen, D., Magaritz, M., Almon, E. & Amiel, A. J. *Water Resources Res.* **23**, 1554–1560 (1987).
14. Mathew, K., Newman, P. W. G. & Ho, G. E. in *Groundwater Recharge with Secondary Sewage Effluents* (Tech. Paper No. 71, Australian Government Publishing Service, Canberra, 1982).
15. Gvirtzman, H., Ronen, D. & Magaritz, M. *J. Hydrol.* **87**, 267–283 (1986).
16. Cohen, Y. & Gordon, L. I. *J. geophys. Res.* **84**, 347–353 (1979).
17. Elkins, J. W., Wofsy, S. C., McElroy, M. B., Kolb, C. E. & Kaplan, W. A. *Nature* **275**, 602–606 (1978).
18. Lipschultz, F. et al. *Nature* **294**, 641–643 (1981).
19. Weiss, R. F. & Price, B. A. *Marine Chem.* **8**, 347–359 (1980).
20. Weiss, R. F. *J. geophys. Res.* **86**, 7185–7195 (1981).
21. Hao, W. M., Wofsy, S. C., McElroy, M. B., Beer, J. M. & Togan, M. A. *J. geophys. Res.* **92**, 3098–3104 (1987).

23. Butler, J. H., Jones, R. D., Garber, J. H. & Gordon, L. I. *Geochim. cosmochim. Acta* **51**, 697–706 (1987).
24. Kaplan, W. A. et al. *Pageoph.* **116**, 423–438 (1978).
25. Yoshinari, D. *Marine Chem.* **4**, 189–202 (1976).
26. Cohen, Y. & Gordon, L. I. *Deep Sea Res.* **25**, 509–524 (1978).
27. Pierotti, D. & Rasmussen, R. A. *Teilus* **32**, 56–72 (1980).
28. Minami, K. & Fukushi, S. *Soil Sci. Pl. Nutr.* **30**, 495–502 (1984).
29. Himmelblau, D. M. *Chem. Rev.* **64**, 527–550 (1964).

Cobalt in ferromanganese crusts as a monitor of hydrothermal discharge on the Pacific sea floor

F. T. Manheim & C. M. Lane-Bostwick

US Geological Survey, Woods Hole, Massachusetts 02543, USA

Ferromanganese oxide crusts, which accumulate on unsedimented surfaces in the open ocean^{1–6}, derive most of their metal content from dissolved and particulate matter in ambient bottom water^{7,8}, in proportions modified by the variable scavenging efficiency of the oxide phase for susceptible ions⁹. They differ in this respect from abyssal nodules, much of whose metals are remobilized from host sediments. Here we present maps of cobalt concentration and inferred accumulation rate of ferromanganese crusts from the Pacific Ocean. We propose that depletion of cobalt in Pacific crusts measures the location and intensity of submarine hydrothermal discharge. Use of the 'cobalt chronometer', an algorithm inversely relating cobalt content and crust growth rate, permits mapping of the accumulation rate of ferromanganese crusts with only indirect recourse to radioactivity-based dating methods. These maps show that crusts in hydrothermal areas grow from two to more than four orders of magnitude faster than in the Central Pacific Ocean. Cobalt-enriched crusts are found where water masses are most isolated from continental-coastal and hydrothermal sources of metals, now and in the past. This relationship can resolve the problem of cobalt enrichment in crusts without recourse to hypotheses invoking special cobalt sources or enrichment mechanisms.

Cobalt, precipitated on MnO₂ substrates as a result of oxidation of Co(II) to Co(III) (refs 10, 11), is the best-defined end member of elements characterizing the hydrogenetic oxide phase. As such, it demonstrates most clearly the phenomena discussed in this letter.

The US Geological Survey (USGS) World Ocean Ferromanganese Crust Database^{12,13} encompasses literature data, especially the Scripps Nodule data bank¹⁴ and its supplement¹⁵. Nearly 1,000 additional samples, many analysed by the US Geological Survey laboratories, have also been incorporated from archived materials and recent cruises (refs 13, 17–23 and data of G. Glasby, personal communication). Crusts and crust-associated (hydrogenetic)¹⁶ nodules are distinguished from true manganese nodules by their encrusting morphology, habitat (presence of hard substrate), preferential occurrence in shallower depth and supportive chemical criteria⁴. To reduce error from contamination of crusts by substrate material and complications arising from the disparity in treatment of hygroscopic and bound water by various authors, we normalized all samples to a 'pure crust' hygroscopic-moisture-free (110 °C overnight) basis, using the Fe + Mn weight as the discriminant and setting the model 'pure crust' at 50 wt% Fe + Mn, close to the highest Fe + Mn value (49.85%) found in the new USGS data set for hydrogenetic crusts¹³. This sample retained about 22% admixture of chemically bound water (H₂O⁺) and other constituents. Thus the normalized per cent by weight of Co, w_{Co}^n , is obtained from the measured weight, w_{Co} , by

$$w_{Co}^n = w_{Co} \times 50 / w_{Fe+Mn}$$

where w_{Fe+Mn} = normalized per cent by weight of Fe + Mn.

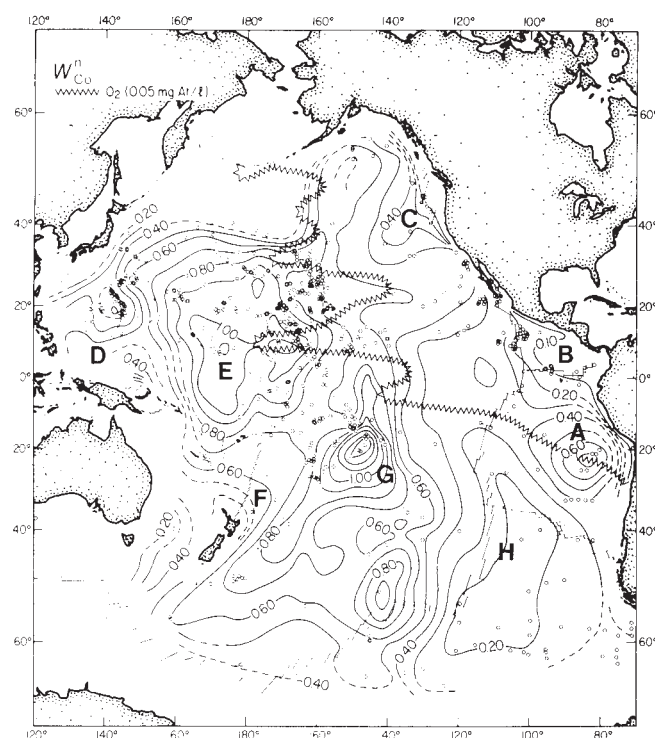


Fig. 1 Normalized per cent cobalt in Pacific crusts (all depths). Contour interval is 0.1. Light lines represent mean location of spreading zones. Jagged lines represent the westward extension of oxygen minimum zone ($O_2 = 0.05 \text{ mg atm l}^{-1}$) (ref. 48). A, Peru basin; B, Galapagos fracture zone; C, Juan de Fuca rift; D, Mariana hydrothermal area; E, Marshall Islands; F, Tonga-Lau-Fiji spreading centre; G, French Polynesia; H, Southeastern Pacific rift.

$w_{\text{Fe}+\text{Mn}}$ was chosen as the normalizing factor because iron and manganese form the building blocks of all crust oxides and because these elements are analysed for nearly all samples. In most cases we used the total thickness of crust, so that differences in thickness (and age) of samples constitute a source of inconsistency.

The samples were computer-gridded and contoured²², with minor manual modifications. The contours, especially in hydrothermal areas, describe weighted average values of variable concentration levels. Procedures are discussed in more detail elsewhere¹³. Contoured maps of cobalt in ferromanganese phases from the Pacific Ocean have been presented previously^{7,23} but these studies did not discriminate between nodules and crusts and were more limited in scope. A recent communication by S. Calvert indicates that the map in ref. 7 excluded shallow samples and is therefore dominated by nodules.

Negative cobalt anomaly zones (<0.3% Co) coincide with known areas of submarine hydrothermal discharge along rift zones associated with the Juan de Fuca, Galapagos and Marianas spreading centres (areas B, C and D in Fig. 1), and spread outward in a halo pattern. Zones low in cobalt are also linked to the southeastern Pacific branch of the rift system, south of the Peru Basin (A), and to the Marianas area. A small halo is found north of the Tonga-Lau-Fiji spreading centre (G). In deposits associated directly with hydrothermal vent systems values lower than 0.01% Co are common, but cobalt content increases rapidly in crust specimens obtained away from the vents²⁴. Similar features have been observed in the Mid-Atlantic Rift zone²⁵. The largest positive cobalt-anomaly area is centred in the Marshall Islands and western Mid-Pacific Mountains (E), and another high-cobalt area with over 1.2% Co is centred in the French Polynesia area (G) at about 20° S.

Because of the well-known inverse relationship between cobalt concentration and water depth of crusts on seamounts^{26,28} one might postulate that the high cobalt-anomaly zones are related to the presence of many shallow seamounts, as in the Mid-Pacific Mountains area. We tested this hypothesis by mapping the distribution of cobalt in crusts from horizons deeper than 4,000 m (Fig. 2). The fact that the cobalt anomaly in the central part of the Pacific is not limited to shallower water depths indicates that it represents a true geographic phenomenon.

Cobalt concentration has been suggested as an indicator of crust accumulation rate²⁸. The inverse logarithmic relationship $R = 6.8 \times 10^{-1} / (w_{\text{Co}}^n)^{1.67}$, where R is rate of accumulation in mm per Myr, has been found to apply in an approximate but surprisingly consistent fashion not only to crusts but also other oxidized manganese-containing phases, including mixed hydrothermal oxides and pelagic sediments⁴. We used this algorithm to map estimated crust accumulation rates from the normalized cobalt concentrations in the Pacific Ocean (Fig. 3). The rates of accumulation vary from <1 mm per Myr in the central parts of the ocean to >1,000 mm per Myr in hydrothermal discharge areas. A hyperbolic formula²⁹ linking cobalt concentration and accumulation rate fits a limited group of Central Pacific crusts better than the above algorithm but is not applicable to hydrothermally-admixed or other more rapidly accreted phases. Other rate algorithms using manganese and iron concentrations have been successfully applied to true abyssal nodule accumulation rates^{30,31}, but are limited to these phases.

The size and distribution of cobalt anomalies in Pacific crusts confirm the expectation that the chemical composition of crusts reflects, at least to a large degree, the concentration of metal sources in the bottom water, integrated over geological time.

Negative cobalt-anomalies in crusts around hydrothermal vent regions in the Pacific Ocean are clarified by recent information on cobalt and other metal content of normal and discharging waters (Table 1). The hydrothermal vent waters are greatly enriched in iron and manganese with respect to normal waters, but Co:Mn and Co:Mn+Fe ratios for both vent fluids and marine hydrothermal oxides are two orders of magnitude lower than in normal ocean waters. Table 1 shows that the Co:(Mn+Fe) ratio is the most similar element of hydrothermal crust and fluid compositions.

Thus, cobalt in crusts is concentrated in inverse proportion to the regional influence of discharging hydrothermal fluids. The greatest negative anomalies (intensity of hydrothermal discharge) occur in the Galapagos fracture zone, consistent with the fact that the thickest and most widespread hydrothermal ferromanganese oxide beds (>15 m thick) found in the oceans to date are located in the Galapagos mounds area³². The next largest region, documented by fewer samples, is the east Pacific

Table 1 Co, Mn and Fe in ocean sediments and waters

Phase	Mn (mg kg^{-1})	Fe (mg kg^{-1})	Co (mg kg^{-1})		Co: (Mn+Fe) $\times 10^4$
			($\times 10^4$)	Co:Mn $\times 10^4$	
Earth's crust	950	41,000	20	210	4.8
Pelagic clay	6,700	58,000	78	116	12.1
Abyssal nodules	186,000	125,000	2,700	145	87
Ocean crusts	216,000	165,000	6,300	291	165
Ocean water $\times 10^3$ (refs 39, 47)	0.044	0.044	0.001	227	114
Marine hydrothermal oxides	273,000	116,000	23	0.84	0.59
Hydrothermal vent waters at 21° N	48	78	0.0078	1.6	0.62

Data sources: solid phases, ref. 4; hydrothermal vent waters from 21° N East Pacific Rise, ref. 46; ocean waters deeper than 500 m from the NE Pacific refs 39, 47.

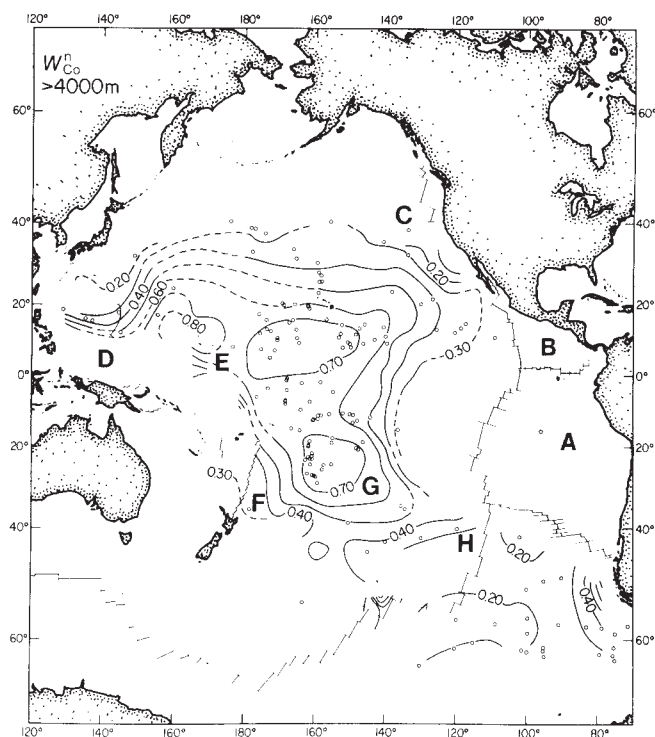


Fig. 2 Normalized cobalt concentration in Pacific crusts, depths >4,000 m. Contour interval in 0.1 per cent cobalt.

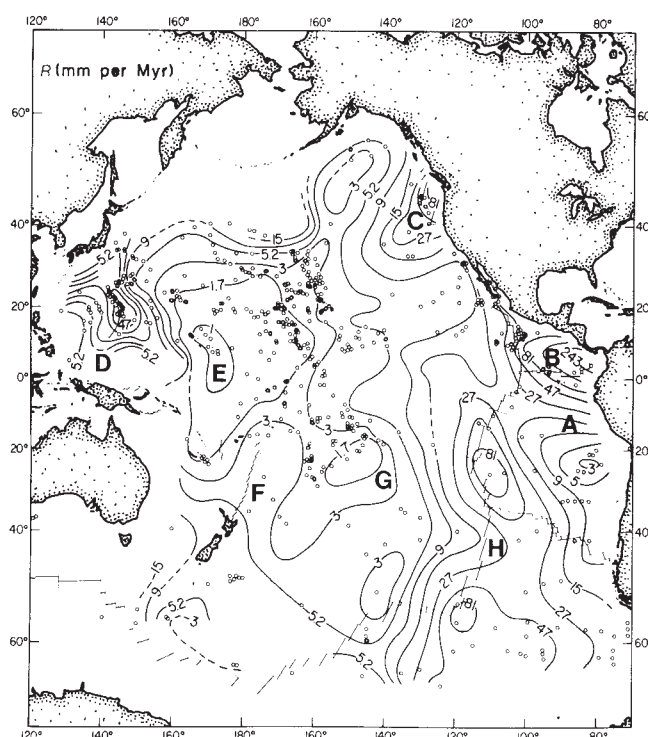


Fig. 3 Crust accumulation rates. Contour interval = 0.5 log₁₀ mm per Myr.

rift zone south of 20° S. Our and other data³³ suggest that this area merits further study.

Boström and Peterson³⁴ first showed that bottom sediments record the flux of hydrothermal manganese and iron. However, sediment composition is affected by such complex factors as variable particulate flux, bottom transport, absorption of exhalates by percolation through sediments, and diagenetic phenomena. The bulk of hydrothermal exhalates is deposited as sediment relatively close to the discharge source³⁵. In contrast the cobalt signal in crusts suggests that crust composition is dominated by hydrothermal sources over much of the eastern Pacific sector, that is, over distances as great as 4,000 km from the spreading centres (see Figs 1 and 3).

The high cobalt (>1.0%) zones occur in central parts of the ocean, far from areas of hydrothermal discharge and terrigenous input, which have been identified, along with biogenic input and atmospheric dust, as sources of manganese and iron³⁶.

The oxygen minimum zone (OMZ) in the Central Pacific has been shown to be accompanied by enhanced cobalt content, possibly supplied in part by horizontal transport of metals from the anoxic slope areas off western North America³⁷⁻³⁹. This was suggested as a key source for cobalt enrichments in Central Pacific crusts^{2,3}, in which case we would expect cobalt content of crusts to be enhanced eastward and to be accompanied by intensification of the oxygen deficit. However, cobalt concentrations generally increase in the opposite direction (Fig. 1) and are high in the French Polynesia area, where the OMZ is poorly developed. Work in progress suggests that water-crust chemical relationships in the OMZ may be complex. At present we assume that the absence of a net regional cobalt crust enrichment associated with the OMZ may be due simply to a compensating supply of manganese and iron.

To summarize, we suggest the relative depletion of bottom water in iron and manganese, rather than specific sources of Co²⁺, accounts for the main features of cobalt enrichment in Pacific crusts. Based on the inverse relationship between cobalt concentration and rate of accumulation, rates of crust accumulation range from <1 mm per Myr in the high-cobalt areas to

>>100 mm per Myr, in the hydrothermal centres. In other words, 10 mm of crust can range in age from ~10³ yr to 10⁷ yr. Because the cobalt contours integrate different crust ages and thicknesses and because of other complicating factors the implications for discharge intensity are only qualitative at this stage. Future palaeoceanographic studies from crusts will benefit from chemical profiles qualified in terms of age equivalents, as has been done for central Pacific sediments⁴⁰. A growing body of 'crust chemical stratigraphic' studies^{19,29,41-44} is now following a pioneering earlier work⁴⁵.

Received 22 February; accepted 21 July 1988.

- Halbach, P., Manheim, F. T. & Otten, P. *Erzmetall* **35**, 447-453 (1982).
- Aplin, A. C. & Cronan, D. S. *Geochim. cosmochim. Acta* **49**, 427-436 (1984).
- Chave, K. E., Morgan, C. L. & Green, W. J. *Appl. Geochem.* **1**, 233-240 (1986).
- Manheim, F. T. *Science* **232**, 601-611 (1986).
- De Carlo, E. H., McMurtry, G. M. & Kim, K. H. *Deep-Sea Res.* **34**, 441-467 (1987).
- Hein, J. R., Morgenson, L. A., Clague, D. A. & Koski, R. A. in *Geology and Resource Potential of the Continental Margin of Western North America and Adjacent Ocean Basins—Beaufort Sea to Baja California* (eds Scholl, D., Grantz, A. & Vedder, J.) 754-771 (Circum-Pacific Council for Energy and Mineral Resources, 1987).
- Calvert, S. E., *Phil. Trans. R. Soc. A* **290**, 43-73 (1978).
- Dymond, J. et al. *Geochim. cosmochim. Acta* **48**, 931-949 (1984).
- Green, W. J. & Chave, K. W. *Mar. Chem.* **23**, 97-108 (1988).
- Burns, R. G. *Geochim. cosmochim. Acta* **38**, 95-102 (1976).
- Murray, J. W. & Dillard, J. G. *Geochim. cosmochim. Acta* **43**, 781-787 (1979).
- Lane, C. M., Manheim, F. T., Hathaway, J. C. & Ling, T. H. US Geological Survey Map MF 1869 (1986).
- World Ocean Database on the Chemical Composition of Ferromanganese Crusts* (eds Manheim, F. T. & Lane-Bostwick, C. M.) (Final report to the US Minerals Management Service, 1987).
- Frazer, J. Z. & Fisk, M. Report prepared for US Department of Interior, Bureau of Mines, SIO Ref. 79-19, 117 (1980).
- National Geophysical Data Center Data Announcement 87-MGG-96 (NOAA, Boulder, Colorado, 1987).
- Halbach, P. *Geol. Rdsch.* **75**, 235-247 (1986).
- Hein, J. R. et al. *US Geological Survey Open File Report No. 85-292*, 129 (1985).
- Hein, J. R., Fleischman, C. L., Morgenson, L. A., Bloomer, S. H. & Stern, R. J. *US Geological Survey Open File Report No. 87-281* 9 (1987).
- Hein, J. R., Schwab, W. C. & Davis, A. S. *Mar. Geol.* **78**, 255-283 (1988).
- De Carlo, E. H., Pennywell, P. A. & Fraley, C. M. *Mar. Min.* **6**, 301-321 (1988).
- Usui, A. *Mar. Geol.* **73**, 311-322 (1986).
- Evenden, G. I. in *Proc. 1986 Workshop Symp. Oceanographic Data Systems* (ed. Stagger, D.) 239-245 (Computer Society Press, Institute of Electrical and Electronic Engineers, Washington, DC, 1986).
- Skornyakova, N. S. & Andryushchenko, P. F. *Int. Geol. Rev.* **6**, 863-919 (1972).

24. Toth, J. R. *Geol. Soc. Am. Bull.* **91**, 44–54 (1980).
25. Goddard, D. A., Thompson, G., Jones, E. J. W. & Okada, H. *Mar. Geol.* **77**, 87–98 (1987).
26. Cronan, D. S. in *Marine Manganese Deposits* (ed. Glasby, G. P.) 11–44 (Elsevier, Amsterdam, 1977).
27. Halbach, P. in *Pacific Mineral Resources: Physical, Economic and Legal Issues* (eds Johnson, C. J. & Clark, A. L.) 137–160 (East-West Mineral Resource Systems Institute, Honolulu, Hawaii, 1986).
28. Halbach, P., Segl, M., Puteanus, D. & Mangini, A. *Nature* **304**, 716–719 (1983).
29. Puteanus, D. *Beitrage zur Meerestechnik* **9**:1, 1–169; **9**:2 1–74 (Technische Universität Clausthal, 1986).
30. Lyle, M. *Geol. Soc. Am. Mem.* **154**, 269–291 (1981).
31. Huh, C.-A. & Ku, T.-L. *Geochim. cosmochim. Acta* **48**, 931–963 (1984).
32. Honnorez, J. *et al. Geol. Soc. Am. Bull.* **92**, 457–472 (1981).
33. Marchig, V. & Gundlach, H. *Bundesanstalt Geowiss. Rohstoffe Circ.* **4**, 3–22 (1987).
34. Boström, K. & Peterson, M. N. A. *Econ. Geol.* **61**, 1258–1265 (1969).
35. Lyle, M. W., Owen, R. M. & Leinen, M. *Init. Rep. DSDP* **92**, 585–596 (1986).
36. Landing, W. M. & Bruland, K. W. *Earth planet. Sci. Lett.* **49**, 45–56 (1980).
37. Knauer, G. A., Martin, J. H. & Gordon, R. M. *Nature* **297**, 49–51 (1982).
38. Jones, C. J. & Murray, J. W. *Limnol. Oceanogr.* **30**:1, 81–92 (1985).
39. Martin, J. H. & Knauer, G. A. *Nature* **314**, 524–526 (1985).
40. Leinen, M. & Stakes, D. *Geol. Soc. Am. Bull.* **90**:1, 357–375 (1979).
41. Halbach, P. & Puteanus, D. *Earth planet. Sci. Lett.* **68**, 73–87 (1984).
42. Halbach, P. & Manheim, F. T. *Mar. Min.* **4**, 319–336 (1984).
43. Von Stackelberg, U., Kunzendorf, H., Marchig, V. & Gwozdz, R. *Geol. Jb.* **75**:A, 213–235 (1984).
44. Olson, D. E., Commeau, R. F. & Manheim, F. T. *Proc. Geol. Soc. Am.* (in the press).
45. Aumento, F. Geological Survey Canada Paper 68-32, 1–30 (1969).
46. Von Damm, K. L. *et al. Geochim. cosmochim. Acta* **49**, 2197–2220 (1985).
47. Gordon, R. M., Martin, J. H. & Knauer, G. A. *Nature* **299**, 611–612 (1982).
48. Brujewicz, S. W. (ed.) *Khimiya Tikhogo Okean* (Izdat. Nauka, Moscow, 1966).

A 1,800-million-year-old Proterozoic gneiss terrane in Islay with implications for the crustal structure and evolution of Britain

F. Marcantonio*, A. P. Dickin*, R. H. McNutt* & L. M. Heaman†

* Department of Geology, McMaster University, Hamilton, Ontario L8S 4M1, Canada

† Department of Geology, Royal Ontario Museum, 100 Queen's Park, Toronto, Ontario M5S 2C6, Canada

A basement gneiss terrane is exposed in the western part of Islay, an island south of Mull in Scotland. This gneiss terrane has been inferred previously^{1–5} to be part of the Lewisian complex (that is, Archean in age), although no direct radiometric information is available. Bentley *et al.*⁶ questioned the Lewisian age for the basement and suggested that the Islay terrane is allochthonous. In this study we employ four different isotope techniques (Rb–Sr, Sm–Nd, Pb–Pb and U–Pb) to determine the crystallization and crustal extraction ages of the Islay gneisses. The isotopic evidence shows that the Islay terrane is early Proterozoic in age and that it is juvenile mantle-derived material rather than a reworking of Archean crust during the Proterozoic. There are, in consequence, two major implications for the crustal evolution of Northern Britain. First, the Grampian terrane, an area directly adjacent to the newly defined Proterozoic Islay block, is probably underlain by Proterozoic basement. Second, Northern Britain can be included in the reconstruction of the Laurentian Shield.

Rock samples were collected from sea cliffs at Portnahaven (PH1 and PH2) and at Port Wemyss (PW1, PW2 and PW3) (Fig. 1); these consist of fine- to medium-grained alkali feldspar, plagioclase, quartz, hornblende, chlorite, biotite and muscovite. PW2 is a highly altered sample and contains very fine-grained sericite replacing the K-feldspar. Epidote is present in all samples, ranging from 2 to 30%. Accessory minerals include calcite, sphene, zircon, apatite and sulphides. The samples have been metamorphosed to amphibolite grade and exhibit a pronounced foliation. Feldspars and hornblende occur commonly as augen enclosed by the foliated micas and chlorite.

Concentrations of Rb and Sr and of Sm and Nd were deter-

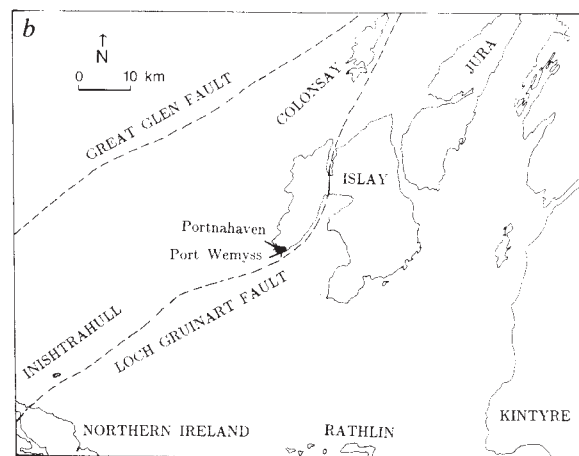
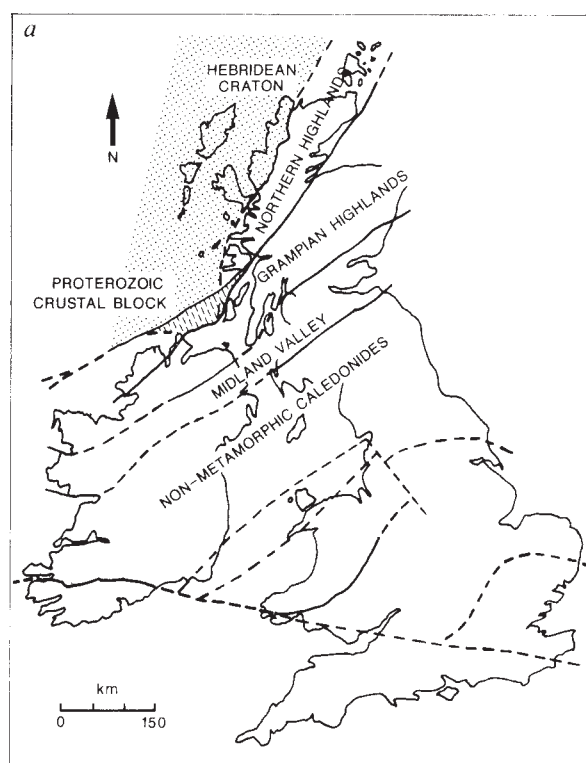


Fig. 1 a, The United Kingdom, showing different tectonic subdivisions. b, Islay and its surrounding area, showing the sample locations.

mined by X-ray fluorescence (XRF) and isotope dilution, respectively. After standard ion-exchange separation⁷ Sr, Sm, Nd and Pb were analysed on a VG 354 mass spectrometer at McMaster University. The zircon analyses on least magnetic abraded zircon fractions from sample PW3 were performed on a VG 354 mass spectrometer at the Royal Ontario Museum using the techniques outlined by Krogh⁸ and Heaman *et al.*⁹

Regression ages from whole rock Rb–Sr, Sm–Nd and Pb–Pb analyses all have large mean squares of weighted deviates (MSWDs). However, the computed ages are in reasonable agreement. The Rb–Sr data define an errorchron with a MSWD of 592 (squares, Fig. 2a), yielding an age of $1,750 \pm 240$ Myr (2σ). The Sm–Nd data yield an age of $1,880 \pm 350$ Myr (Fig. 2b). The large uncertainty in the Sm–Nd age partially reflects the small variation in Sm/Nd ratios. On a Pb–Pb diagram (Fig. 2c) the samples define a slope of 0.105 ± 0.007 (2σ), which corresponds

Table 1 Geochemical and isotopic data on the Islay gneisses

Whole rock	PH1	PH2	PW1	PW2	PW3
Rb (p.p.m.)*	58	112	32	7	49
Sr (p.p.m.)	598	1,315	261	1,926	249
$^{87}\text{Sr}/^{86}\text{Sr}^\dagger$	0.70960	0.70812	0.71310	0.70308	0.71683
$^{87}\text{Rb}/^{86}\text{Sr}$	0.2832	0.2475	0.3556	0.0116	0.5705
$\epsilon_{\text{Sr}(t=1,780)}^\ddagger$	-0.83	-8.90	22.63	5.33	-2.58
Sm (p.p.m.)§	7.1	14.9	13.1	4.0	
Nd (p.p.m.)	39.0	98.0	46.9	68.7	23.1
$^{143}\text{Nd}/^{144}\text{Nd} $	0.51171	0.51150	0.51161	0.51179	0.51164
$^{147}\text{Sm}/^{144}\text{Nd}$	0.1100	0.0917	0.1031	0.1148	0.1058
$\epsilon_{\text{Nd}(t=1,780)}^\P$	1.71	1.79	1.33	2.17	1.30
$T_{(\text{DM})}^\#$	1,959	1,927	1,975	1,930	1,982
$^{206}\text{Pb}/^{204}\text{Pb}^{**}$	16.478	17.965	18.766	17.440	16.821
$^{207}\text{Pb}/^{204}\text{Pb}$	15.289	15.489	15.531	15.412	15.356
$^{208}\text{Pb}/^{204}\text{Pb}$	35.591	36.564	36.524	36.574	35.461

PW3 Zircon analyses (from the least magnetic fraction)			
	Best abraded: pink	Best abraded: brown cloudy with cracks	2nd best abraded: pink with cracks
$^{206}\text{Pb}/^{238}\text{U}^\ddagger$	0.3138	0.3050	0.3133
$^{207}\text{Pb}/^{235}\text{U}$	4.7047	4.5854	4.7118
$^{207}\text{Pb}/^{206}\text{Pb}$	0.1088	0.1090	0.1091
$^{206}\text{Pb}/^{238}\text{U}$ age	1,759 Myr	1,716 Myr	1,757 Myr
$^{207}\text{Pb}/^{235}\text{U}$ age	1,768 Myr	1,747 Myr	1,769 Myr
$^{207}\text{Pb}/^{206}\text{Pb}$ age	1,779 Myr	1,783 Myr	1,784 Myr
U (p.p.m.) ‡‡	250	340	203
Pb (p.p.m.)	83	112	65

* Precision on Rb and Sr concentrations by XRF is 1.0% (2 σ). † Precision on Sr isotopic ratios is 0.002% (2 σ). Average obtained for $^{87}\text{Sr}/^{86}\text{Sr}$ on 126 analyses of NBS 987 is 0.710233. ‡ ϵ_{Sr} is measured in parts per ten thousand relative to Bulk Earth at 1,780 Myr. § Precision on Sm and Nd concentrations by isotope dilution is 0.2% (2 σ). $||$ Precision on Nd isotopic ratios is 0.002% (2 σ). Average obtained for $^{143}\text{Nd}/^{144}\text{Nd}$ on 75 analyses of La Jolla is 0.511853. ¶ ϵ_{Nd} is measured in parts per ten thousand relative to BE at 1,780 Myr. $^\#$ Model ages based on extraction from a depleted mantle reservoir¹². ** Pb whole rock isotope analyses are within 0.2% (2 σ). ‡‡ U/Pb ratios are within 0.25% (2 σ) and $^{207}\text{Pb}/^{206}\text{Pb}$ ratios are within 0.05% (2 σ). ‡‡ Precision on U and Pb concentrations by isotope dilution is 2% (2 σ).

to an age of $1,720 \pm 244$ Myr (MSWD = 11.6). A single-stage μ value ($^{238}\text{U}/^{204}\text{Pb}$ ratio) of 7.9 was obtained for the source of the gneisses. Three U-Pb zircon analyses define a discordia line (Fig. 3) with an upper intercept age of $1,782 \pm 5$ Myr (the $^{207}\text{Pb}/^{206}\text{Pb}$ ages are 1,779, 1,783 and 1,784 Myr; see Table 1).

The precise U-Pb zircon age of 1,782 Myr was used to calculate initial Nd and Sr isotopic ratios. The $^{87}\text{Sr}/^{86}\text{Sr}$ and $^{143}\text{Nd}/^{144}\text{Nd}$ ratios of Bulk Earth¹⁰ at that time were 0.70241 and 0.51033 respectively. The $\epsilon_{\text{Nd}(t)}$ values¹¹ (average of +1.7) suggest derivation from a depleted mantle source. The $\epsilon_{\text{Sr}(t)}$ values vary from a depleted signature of -9.6 to an enriched signature of +21.5. This variation probably represents open-system behaviour of the Rb-Sr isotopic system during Caledonian metamorphism. This would also account for the large MSWD and error associated with the Rb-Sr whole rock errorchron.

Assuming extraction from a depleted mantle reservoir, model Sm-Nd ages¹², $T_{(\text{DM})}$, average at 1,955 Myr. Discrepancy between the $T_{(\text{DM})}$ model ages and the U-Pb zircon age can be explained in two ways: (1) the zircon age denotes the time of crustal extraction and the slightly older $T_{(\text{DM})}$ age indicates that a small Archean component may be involved, or (2) the zircon age represents remelting, while the $T_{(\text{DM})}$ age represents the true time of crustal extraction. In the latter case one expects inheritance, which is not seen in the zircon fractions because there are no core overgrowths.

The high model μ_1 value of 7.9 documented for the Islay gneisses precludes significant Lewisian crustal involvement, since the Lewisian crust¹³ is depleted in U relative to Pb and hence has a lower μ value. Furthermore, normal mantle μ_1 values approach 8 for rocks of Proterozoic age¹⁴. Hence, the initial Nd ratios, the model Nd ages and the model μ value all

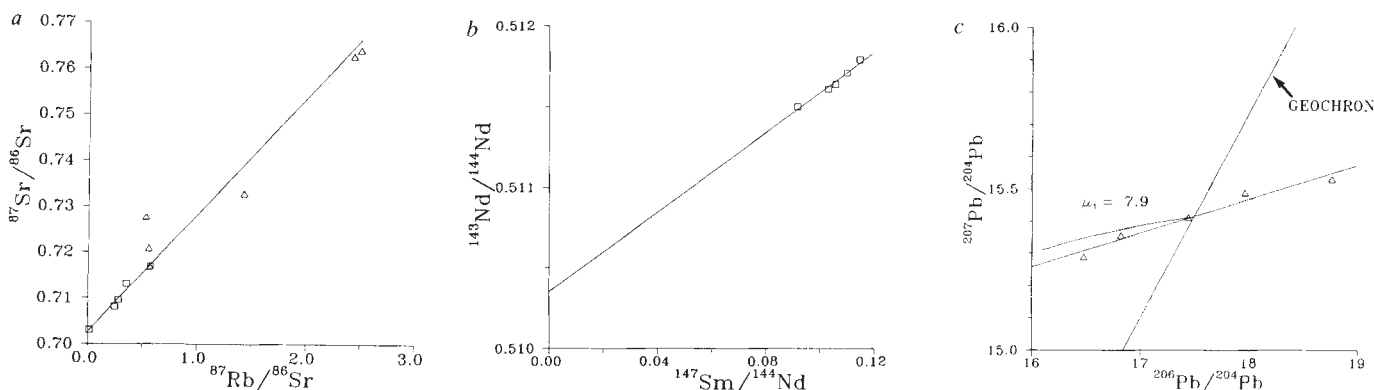


Fig. 2 Isochron plots of *a*, $^{87}\text{Sr}/^{86}\text{Sr}$ versus $^{87}\text{Rb}/^{86}\text{Sr}$ (data on Islay gneisses are plotted as squares and the data of MacIntyre and van Breemen⁵ are plotted as triangles). *b*, $^{143}\text{Nd}/^{144}\text{Nd}$ versus $^{147}\text{Sm}/^{144}\text{Nd}$ *c*, $^{207}\text{Pb}/^{204}\text{Pb}$ versus $^{206}\text{Pb}/^{204}\text{Pb}$.

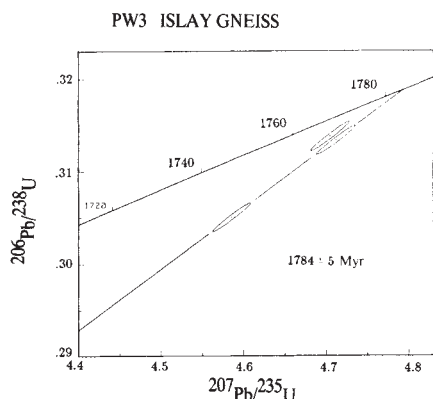


Fig. 3 U-Pb concordia diagram for zircon fractions from PW3.

imply that the Islay gneisses constitute a Proterozoic mantle-derived terrane. The terrane is not simply a reworking of Archean crustal material during the Laxfordian metamorphic event, as seen elsewhere in Scotland¹⁵.

Using magnetic, gravity and seismic geophysical data, Evans *et al.*⁴ have constructed a structural map (Fig. 1b) of the offshore region surrounding Islay. The area from Northern Ireland to the tip of Colonsay, the Great Glen fault and its splay, the Loch Gruinart fault, define a wedge-shaped sliver. Using seismic reflection data Brewer *et al.*¹⁶ have confirmed the existence of this crustal block by tracing the outline of the Great Glen and the Loch Gruinart faults. It is this Colonsay-western Islay terrane which Bentley *et al.*⁶ proposed to be allochthonous. The island of Inishtrahull (Fig. 1b), where a basement gneiss complex is also found, is included in their⁶ definition of the terrane since the rocks there are similar to the Islay gneisses⁵. Inishtrahull samples⁵ are plotted as triangles on Fig. 2a; these show a close isotopic affinity to our data from Islay. After excluding the most aberrant whole rock point, a composite regression age of $1,775 \pm 240$ Myr is obtained from the Islay and Inishtrahull samples. Given that both belong to the same crustal block⁴, and in light of the model Nd age data for Islay, we suggest that the rocks of Inishtrahull are probably Proterozoic in age.

Magnetic evidence¹⁷ suggests that basement continuity exists between the Islay terrane and the adjacent Grampian Highlands. Pidgeon and Aftalion¹⁸ have proposed that Proterozoic crust may be the source for the Caledonian granites of the Grampian block. Their U-Pb zircon analyses¹⁸ yielded upper intercept ages of ~ 1600 Myr and suggest that the Highland Boundary Fault marks the southern extent of a 'probable' Proterozoic basement. Their upper-intercept ages, somewhat younger than Islay, may indicate complex Pb loss patterns from the unabraded bulk zircon fractions used in their study. Furthermore, an average $T_{(DM)}$ model age of 1,600 Myr is obtained from Sm-Nd data^{19,20} for the Caledonian granites within the Grampian block. This average model age may have been biased downwards, in comparison with our value of 1,955 Myr by mixing between Proterozoic basement and Caledonian mantle-derived magma.

The newly defined Proterozoic Belt in Scotland has affinities with areas outside Britain including the Makkovik, Ketilidian and Svecofennian crustal belts of the Laurentian craton²¹⁻²⁴. For instance, Patchett and Bridgwater²¹ have argued that granites in the Ketilidian mobile belt of Greenland were derived in the Proterozoic, with 10% reworking of older basement. Patchett and Kouvo²² proposed that the 1.9–1.7 Gyr crust of the Svecofennian terrane of South Finland is derived by mantle differentiation, again with minor reprocessing of Archean crust. We suggest that the gneiss terrane on Islay is the missing crustal segment between the Ketilidian and Svecofennian crustal zones and completes the Proterozoic orogenic belt of the Laurentian margin.

We acknowledge M. Dickin and M. Hardie for assistance with fieldwork and thank P. J. Patchett for helpful comments. F.M.

gratefully acknowledges NSERC for scholarship support during postgraduate studies. Isotopic research at McMaster is supported by NSERC grants to A.P.D. and R.H.M. Geochronological research at the Royal Ontario Museum is supported by an NSERC grant to T. Krogh.

Received 15 March; accepted 14 July 1988.

1. Johnstone, G. S. in *British Regional Geology—The Grampian Highlands* 3rd Edn (Institute of Geological Sciences, 1966).
2. Stewart, A. D. *Mem. Am. Ass. Petrol. Geol.* **12**, 595–608 (1969).
3. Stewart, A. D. *Geol. Soc. Lond. spec. Rep.* **6**, 43–51 (1975).
4. Evans, D., Kenolty, N., Dobson, M. R. & Whittington, R. J. *Report 79/15*, Institute of Geological Sciences (Natural Environment Research Council, 1980).
5. MacIntyre, R. M. & Van Breemen, O. *Sci. Proc. Royal Dublin Soc.* **A5**, (19), 301–309 (1975).
6. Bentley, M. R., Maltman, A. J. & Fitches, W. R. *Geology* **16**, 26–28 (1988).
7. Halliday, A. N. *et al. Earth planet. Sci. Lett.* **63**, 241–256 (1983).
8. Krogh, T. E. *Geochim. cosmochim. Acta* **46**, 637–649 (1982).
9. Heaman, L. M., Machado, N., Krogh, T. E. & Weber, W. *Contr. Miner. Petrol.* **94**, 1, 82–89 (1986).
10. Allègre, C. J. *Tectonophysics* **81**, 109–132 (1982).
11. DePaolo, D. J. & Wasserburg, G. J. *geophys. Res. Lett.* **3**, 249–252 (1976).
12. DePaolo, D. J. *Eos* **62**, no. 14, 137–140 (1981).
13. Chapman, H. J. & Moorbath, S. *Nature* **268**, 41–42 (1977).
14. Moorbath, S. & Taylor, P. N. in *Precambrian Plate Tectonics* (ed. Kröner, A.) 419–525 (Elsevier, Amsterdam, 1981).
15. Taylor, P. N., Jones, N. W. & Moorbath, S. *Phil. Trans. R. Soc. A310*, 605–625 (1984).
16. Brewer, J. A. *et al. Nature* **305**, 206–210 (1983).
17. Westbrook, G. K. & Borradaile, G. J. *Scott. J. Geol.* **14**, no. 3, 213–224 (1978).
18. Pidgeon, R. & Aftalion, M. *Geol. J. spec. Issue* **10**, 183–220 (1978).
19. Halliday, A. *Nature* **307**, 229–233 (1984).
20. Hamilton, P. J., O'Nions, R. K. & Pankhurst, R. J. *Nature* **287**, 279–284 (1980).
21. Patchett, P. J. & Bridgwater, D. *Contr. Miner. Petrol.* **87**, 311–318 (1984).
22. Patchett, P. J. & Kouvo, O. *Contr. Miner. Petrol.* **91**, 1–12 (1986).
23. Gower, C. F. & Owen, V. *Can. J. Earth Sci.* **21**, 678–693 (1984).
24. Kalsbeek, F. & Taylor, P. N. in *The Deep Proterozoic Crust in the North Atlantic Provinces* (eds Tobi, A. C. & Touret, J. L. R.) 237–245 (Reidel, Dordrecht, 1986).

The hyporheic habitat of river ecosystems

J. A. Stanford* & J. V. Ward†

* Flathead Lake Biological Station, University of Montana, Polson, Montana 59860, USA

† Department of Biology, Colorado State University, Fort Collins, Colorado 80523, USA

Contemporary river ecology is based primarily on biogeochemical studies of the river channel and interactions with shoreline vegetation, even though most rivers have extensive floodplain aquifers that are hydraulically connected to the channel. The hyporheic zone, the interstitial habitat penetrated by riverine animals, is characterized as being spatially limited to no more than a few metres, in most cases centimetres, away from the river channel¹⁻⁹. However, riverine invertebrates were collected in hundreds per sample within a grid of shallow (10 m) wells located on the floodplain up to 2 km from the channel of the Flathead River, Montana, USA. Preliminary mass transport calculations indicate that nutrients discharged from the hyporheic zone may be crucial to biotic productivity in the river channel. The strength and spatial magnitude of these interactions demonstrate an unexplored dimension in the ecology of gravel-bed rivers.

The Flathead River, a major tributary of the Clark Fork of the Columbia, drains a catchment area of 22,241 km² (average flow = 340 m³ s⁻¹) in northwestern Montana and southeastern British Columbia. The morphology and surficial geology of the Kalispell Valley (Fig. 1) were largely determined by Pleistocene glaciation. Fine-grained lacustrine sediments and fault traces delineate the northern geological boundary of the palaeodeltas of proglacial Flathead Lake. The river is heavily braided in the area delineating this boundary or ectone (E in Fig. 1). An alluvium of cobbles, gravel and sand covers an impermeable clay formation of Tertiary age upstream of the heavily-braided reach.

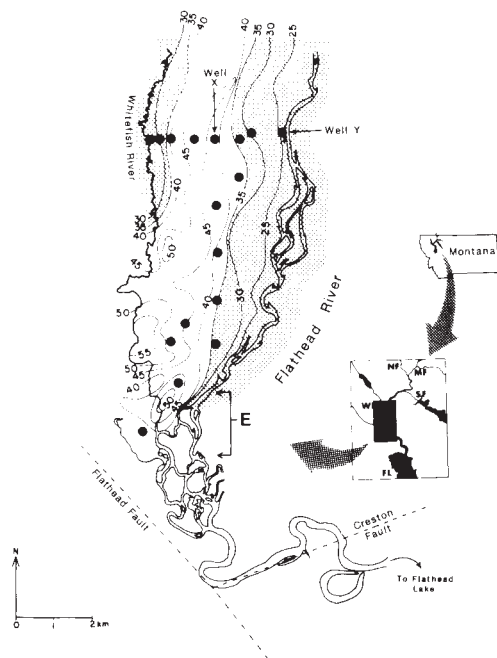


Fig. 1 Map of hyporheic habitat (stippled areas) within the Kalispell Valley of the Flathead River. Isopleths of specific conductance (mS per metre) were determined from 271 wells. Closed circles locate only those wells from which biological samples were obtained. Inset map shows the three major tributaries upstream of the study site. NF, North Fork; MF, Middle Fork; SF, South Fork; the Whitefish River (WF) and Flathead Lake (FL). The area designated E is the interface between porous and less transmissive substrata (see text).

Groundwaters in this alluvium interact hydraulically with the Flathead River and a tributary system, the Whitefish River (Fig. 1). Wells near the river channel produced hydrographs that closely tracked daily and seasonal changes in river flow (Fig. 2). Even wells in the centre of the valley, 2 km or more from the rivers, were influenced by river-flow patterns (Fig. 2). The aquifer lies on a 2° slope, bounded on both sides by the river channels; water flows through the system in a north to south

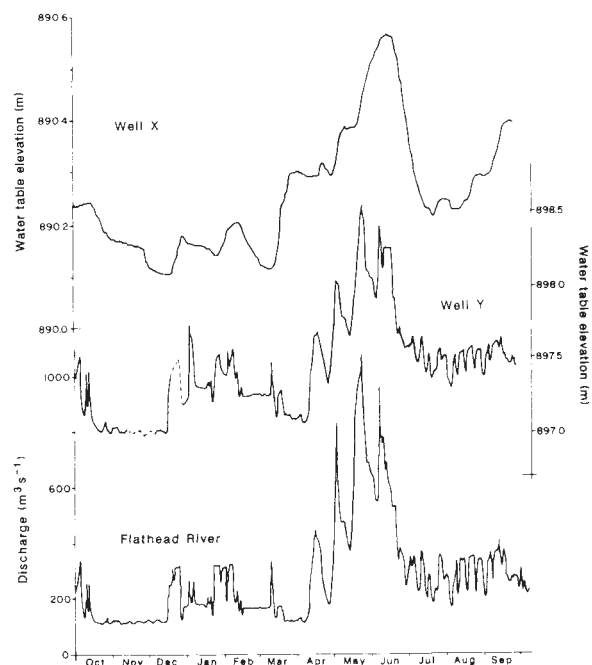


Fig. 2 Hydrographs from wells X and Y (see Fig. 1) compared with the Flathead River in 1984-85. Nonseasonal fluctuations were caused by discharges from a large hydroelectric dam on an upstream tributary (SF in Fig. 1). Well X is 2.4 km and Well Y 50 m from the Flathead River channel.

direction at an average rate of $0.7 \text{ m}^3 \text{ s}^{-1}$. The primary area of aquifer discharge is near the confluence of the two rivers, where faulting and finer sediments on the proglacial Flathead Lake palaeodeltas begin to limit groundwater flow rates (E in Fig. 1). Water also moves to and from the channels. During spring freshet, water moves into the aquifer from the river channel until a hydrological equilibrium occurs or river flow begins to decrease. As flows decrease, aquifer water is discharged from the hyporheic zone into the river.

Spatial plots of specific conductance data (Fig. 1) also demonstrated lateral dilution. The 35 mS per metre isopleth (Fig. 1) appeared to delineate true interstitial groundwaters that are inhabited by subterranean fauna¹⁰⁻¹⁴ and are less interactive with surface waters.

This was confirmed by the distribution and abundance patterns of the interstitial fauna within the aquifer. Biota collected from wells in the hyporheic zone (as defined by the 35 mS per metre isopleth, Fig. 1), consisted almost exclusively of stonefly (Plecoptera) larvae and other typically riverine taxa (Table 1). In contrast, subterranean forms were abundant in most of the wells located on the high concentration side of the 35 mS per metre isopleth. There was some overlap between riverine and subterranean taxa in several wells located near the 35 isopleth, but either subterranean amphipods (*Stygobromus* spp.) or stonefly larvae were very abundant (hundreds per sample), never both.

Levels of dissolved oxygen within the wells were always greater than 50% saturation. Temperatures remained at 7-9 °C throughout the year in all wells, except those located near the river channel where greater exchange of water elevated or reduced temperatures in a seasonal way. It may be that the biota navigate through the interstitial groundwaters by following thermal cues when near the river channel, or ion concentration gradients when far from the river.

On the basis of the relative distribution of the biota and the 35 mS per metre isopleth, we were able to measure the hyporheic habitat within the 10 km (approximately) study area (Fig. 1). The hyporheic zone is on average about 3 km wide with an average depth of about 10 m; whereas, the Flathead River chan-

Table 1 Relative abundance of biota

Hyporheic zone	
<i>Paraperla frontalis</i> (Insecta: Plecoptera)	100
<i>Isocapnia</i> 4 spp. (Insecta: Plecoptera)	100
Chironomidae (Insecta: Diptera)	10
Capniidae (Insecta: Plecoptera)	20
Early instar zoobenthos	20
Incidental zoobenthos	20
Phreatic zone*	
<i>Stygobromus</i> 2 spp. (Crustacea: Amphipoda)	100
Cyclopoid copepods (Crustacea: Eucopopoda)	100
<i>Asellus</i> sp. (Crustacea: Isopoda)	10
Bathynellidae (Crustacea: Bathynellacea)	10

Biota were collected from 17 wells located within the alluvial aquifer adjacent to the Flathead River and on either side of the 35 mS per metre specific conductance isopleth, which differentiated phreatic (>35 mS per metre) and hyporheic (<35 mS per metre) habitats. Data are the percentage of total wells located in either the phreatic or hyporheic zones in which a particular taxon was present in numbers greater than 10 per well on every sampling date ($n=8$) in 1984-6.

* All taxa listed under phreatic zone are crustaceans of subterranean facies. Bathynellids and *Stygobromus* are exclusively subterranean groups. *Asellus* (*Caecidotea*) sp. is a blind, depigmented isopod of attenuated morphology. Although most cyclopoid copepods are planktonic or littoral, several species are widespread in interstitial biotopes¹⁰⁻¹⁴.

Table 2 Average concentrations and standard deviations of nutrients at well sites

Sampling site	Soluble reactive phosphorus	Soluble total phosphorus	Nitrite + nitrate nitrogen	Soluble organic carbon
Well X	1.5 ± 0.5	4.9 ± 1.0	916.0 ± 11.0	1,810 ± 138
Well Y	1.2 ± 0.5	2.2 ± 0.8	145.0 ± 5.0	1,810 ± 112
River	BDL*	2.3 ± 0.4	38.0 ± 1.0	1,730 ± 162

Concentrations are in μg per litre. Well sites were located 2.5 km (well X) and 50 m (well Y) from the Flathead River channel compared to values in the river. These data summarize six bimonthly sampling dates during 1984–85.

* Below detection limit = $1.0 \mu\text{g l}^{-1}$.

nel (median flow) is about 50 m wide with most zoobenthos in the upper 0.25 m of channel substrata. There is therefore about 0.3 km^3 of hyporheic habitat, compared to about $125,000 \text{ m}^3$ of channel habitat. Standing crop biomass in the hyporheic zone could easily exceed benthic biomass in this river.

Stoneflies are apparently the top consumers within an as yet undescribed and probably detritus-based food chain. Others^{4,15} have suggested that the hyporheic zone is a functional sink for fine ($<500 \mu\text{m}$) particulate organic detritus from the channel. Particulates may also be recruited by infiltration of precipitation through the soil profile. Decomposition of organic detritus, coupled with ionic leaching, desorption and other biophysical processes, like nitrification, may sequester labile (bioavailable¹⁶) nutrients within groundwaters. Nutrient concentrations were significantly higher in the hyporheic (for example well X, Table 2) than in the river. Preliminary calculations of mass transport of bioavailable phosphorus¹⁶ and nitrate–nitrogen indicate that baseflow loads in the Flathead River increased by 25% and 12% respectively. During extended baseflow periods, certain near-shore areas of the relatively unproductive Flathead River channel are matted with algae. We believe that these areas of enhanced phytobenthos production occur in direct response to the inflow of nutrient-rich hyporheic waters.

Gravel-bed rivers are common worldwide¹⁷. The practice of screening groundwater monitoring wells accounts for the absence of faunal records in samples collected by hydrogeologists. The spatial extent and strength of hyporheic-channel interactions undoubtedly vary from river to river. Nonetheless, hyporheic-channel interactions as reported here are probably common features of gravel-bed river segments, and should be included in holistic constructs¹⁸ of riverine ecosystems.

We thank R. Noble (Montana Bureau of Mines and Geology) for his help. This research was supported by a National Science Foundation (USA) grant and the Montana Department of Natural Resources and Conservation.

Received 19 January; accepted 11 July 1988.

- Stanford, J. A. & Gaulin, A. R. *Science* **185**, 700–702 (1974).
- Schwoerbel, J. *Arch. Hydrobiol. Suppl.* **33**, 1–62 (1967).
- Husmann, S. *Smithson. Contr. Zool.* **76**, 161–169 (1971).
- Williams, D. D. & Hynes, H. B. N. *Freshwater Biol.* **4**, 233–256 (1974).
- Danielopol, D. L. *Int. J. Speleol.* **8**, 23–51 (1976).
- Danielopol, D. L. *Int. Rev. ges Hydrobiol.* **65**, 777–791 (1980).
- Bretschko, G. *Verh. int. Verein. Limnol.* **22**, 2049–2052 (1985).
- Williams, D. D. in *The Ecology of Aquatic Insects* (eds Resh, V. H. & Rosenberg, D. M.) 430–455 (Praeger, New York, 1984).
- Pennak, R. W. & Ward, J. V. *Arch. Hydrobiol. Suppl.* **74**, 356–396 (1986).
- Holsinger, J. R. *Am. J. Zool.* **74**, 146–153 (1988).
- Ward, J. V. *Trans. Am. microsc. Soc.* **96**, 452–466 (1977).
- Holsinger, J. R. *Crustaceana Suppl.* **4**, 244–281 (1977).
- Ward, J. V. & Holsinger, J. R. *Int. J. Speleol.* **11**, 63–70 (1981).
- Pennak, R. W. & Ward, J. V. *Trans. Am. microsc. Soc.* **104**, 209–222 (1985).
- Hynes, H. B. N. *Hydrobiologia* **100**, 93–99 (1983).
- Ellis, B. K. & Stanford, J. A. *Bioavailability of Phosphorus Fractions in Flathead Lake and its Tributary Waters* (Open File Rpt. 81, Flathead Lake Biology Station, University of Montana, Polson, 1986).
- Sediment Transport in Gravel Bed Rivers* (eds Thorne, C. R., Bathurst, J. C. & Hey, R. D.) (Wiley, Chichester, 1987).
- Vannote, R. A., Minshall, G. W., Cummins, K. W., Sedell, J. R. & Cushing, C. E. *Can. J. Fish. aquat. Sci.* **37**, 130–137 (1980).

Embryonic acetylcholine receptors guarantee spontaneous contractions in rat developing muscle

Fernán Jaramillo, Stefano Vicini* & Stephen M. Schuetze

Department of Biological Sciences, Columbia University, New York 10027, USA

*Fidia-Georgetown Institute, 3900 Reservoir Road NW, Washington, DC 20007, USA

Many proteins are expressed in distinct embryonic and adult forms^{1,2}. However, in most cases we do not know why the embryonic form of proteins is required. This question can be readily addressed for the acetylcholine receptor (AChR) because developmentally specified modifications of this ligand-gated ion channel can be directly related to changes in membrane currents. In developing rat soleus muscle, spontaneous transmitter release causes miniature end-plate currents (m.e.p.cs) to flow into the muscle cell. We show here that these m.e.p.cs in neonatal soleus trigger spontaneous contractions. By injecting m.e.p.cs into young fibres, we showed that only embryonic m.e.p.cs can trigger such contractions; adult m.e.p.cs do not last long enough. Developing muscle fibres must be active for synapse and muscle differentiation. Our experiments indicate that the embryonic form of the AChR is essential for spontaneous contractile activity and may therefore be required for normal neuromuscular development.

Neonatal rat soleus fibres were observed to contract in response to spontaneous muscle action potentials. Contraction rates in excess of five per minute at 37°C were common. In almost every case, action potentials were triggered by spontaneous miniature end-plate potentials (m.e.p.ps) (Fig. 1). These are changes in the potential across the muscle membrane due to m.e.p.cs caused by the spontaneous release of packets of acetylcholine from the nerve. In some cells, nearly every m.e.p.p. caused an action potential and in 8 of 55 cells more than 50% of all m.e.p.ps triggered action potentials. The average percentage of super-threshold m.e.p.ps was $19 \pm 27\%$ (mean $\pm$ standard deviation; 10–76 m.e.p.ps per cell, ages 2–8-day-old). During the first three postnatal weeks the average time required for a m.e.p.p. to decay to half its peak amplitude was reduced from $8.9 \pm 2.4 \text{ ms}$ ($n=9$ cells) to $2.4 \pm 0.5 \text{ ms}$ ($n=10$). During this period, m.e.p.p. amplitudes declined from $11.1 \pm 3.1 \text{ mV}$ in days 1–3 ($n=21$) to $1.28 \pm 0.26 \text{ mV}$ in days 19–22 ($n=10$). By two weeks after birth no m.e.p.ps caused action potentials and no spontaneous contractions were observed. It is unlikely that the large neonatal m.e.p.p. amplitudes are due to the simultaneous release of two or more quanta of acetylcholine because histograms of m.e.p.p. and m.e.p.c. amplitudes appeared unimodal ($n=40$, 10–250 m.e.p.ps/m.e.p.cs per cell).

Adult-type AChRs gradually replace the embryonic type during the first three weeks after birth. The time-course^{3–6} and molecular basis⁷ of this developmental process have been characterized in detail. Adult-type AChRs have an apparent mean channel open time which is 25% of the embryonic one ($\sim 1.5 \text{ ms}$ at room temperature) and an $\sim 50\%$ larger conductance (50 pS) than the embryonic-type. These features account for the shorter duration and larger amplitude of adult m.e.p.cs. To estimate how changes in the amplitude and time course of m.e.p.cs affect the amplitude and time course of m.e.p.ps, we injected m.e.p.cs with an intracellular microelectrode at synaptic sites and recorded the resulting 'pseudo-m.e.p.ps' with an independent microelectrode (Fig. 2). Muscle cells responded similarly to spontaneous m.e.p.cs and to currents injected electronically (Fig. 2a). We could therefore directly compare the response of muscle cells to embryonic or adult m.e.p.cs (Fig.

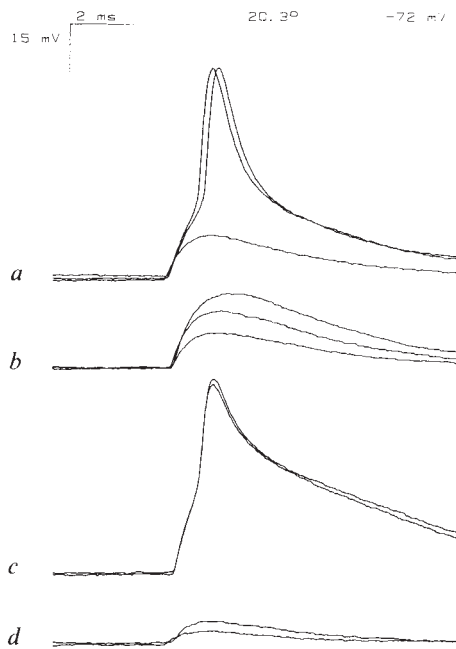


Fig. 1 Pharmacology of spontaneous events recorded from a soleus end-plate four days after birth. The rising phases of the events were aligned to facilitate comparison. *a*, Events recorded in medium alone. *b*, 5 min after $10 \mu\text{M}$ tetrodotoxin (TTX, Sigma) was added to the bath. *c*, 10 min after TTX was washed out. *d*, 1 min after $1 \mu\text{g ml}^{-1}$ α -BGT (Sigma) was added to the bath. Spontaneous events disappeared completely, shortly after those shown in *d* were recorded. The membrane potential was -72 mV . Action potentials triggered by m.e.p.s were also observed at 37°C . **Methods.** Soleus muscles were removed from a 4-day-old rat (Sprague-Dawley; Charles River), bathed in L-15 (4.2 mM Ca^{2+}) tissue-cultured medium (Grand Island Biological), and prepared for electrophysiology as described previously⁶. End-plates were localized visually⁶ and intracellular recordings were made within $10\text{--}50 \mu\text{m}$ of the end-plate. Electrodes were filled with 3 M KCl .

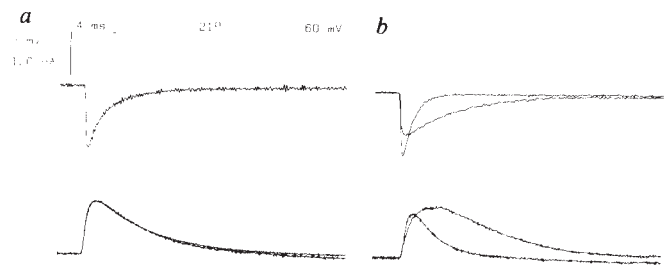


Fig. 2 Injections of m.e.p.s into an 11-day-old fibre. *a*, Average of 25 m.e.p.s (upper) recorded with a two-electrode voltage clamp (-60 mV). Average of 25 m.e.p.s (lower) recorded after voltage clamp (resting potential was -60 mV). Superimposed (lower) is the average response to 25 injections (under non voltage clamp conditions) of the recorded m.e.p.c. In other cells true m.e.p.s had $(8 \pm 13)\%$ ($n=6$) larger amplitudes than pseudo-m.e.p.s. *b*, (upper) Injections of typical embryonic ($\tau=5 \text{ ms}$) and adult ($\tau=1 \text{ ms}$) m.e.p.s into the cell mentioned in *a*. $5 \mu\text{M}$ TTX was present in the bath. 'Typical' embryonic and adult m.e.p.s used for current injections were the average of 24 and 40 m.e.p.s recorded in a 1-day-old fibre and a 24-day-old fibre respectively. The adult m.e.p.c. amplitude was set to match the average amplitude of spontaneous m.e.p.s in this fibre (spontaneous m.e.p.s increased from $\sim 2.6 \text{ nA}$ at day 1 to $\sim 3.6 \text{ nA}$ at day 22, $n=13$). The embryonic m.e.p.c. amplitude was set at 66% of the adult value to allow for the developmental increase in m.e.p.c. amplitude. *b*, (lower) Embryonic pseudo-m.e.p.s had $(14 \pm 6)\%$ ($n=5$) larger amplitudes and ~ 2.5 -times longer durations than adult pseudo-m.e.p.s. **Methods.** voltage clamp recordings and current injections were performed with an Axoclamp 2A (Axon Instruments) whose current and voltage outputs were fed into an LSI 11/73 computer (Digital Equipment Corporation). M.e.p.s were filtered at $2\text{--}5 \text{ kHz}$ (-3 dB , 8 pole), sampled at 20 kHz and averaged as described by Land and colleagues¹⁵. High ($30\text{--}90 \text{ M}\Omega$) electrode resistances were required to prevent cell damage in young fibres. The $20\text{--}80\%$ rise time to a 2 mV hyperpolarizing pulse was $69 \pm 22 \mu\text{s}$ ($n=16$) and the average m.e.p.p. amplitude under voltage clamp was less than 10% of the average m.e.p.p. in the unclamped fibre. Before current injections, a 1 nA current pulse was passed into the bath and the electrode capacitance was electronically neutralized until the $20\text{--}80\%$ rise time of the voltage drop across the electrode was $25\text{--}50 \mu\text{s}$.

2*b*). A major advantage of this approach is that it excludes other factors, such as fibre size and specific membrane properties, that also affect the amplitude and time course of m.e.p.s. Injection experiments revealed that the long duration of m.e.p.s in young fibres is largely due to the long duration of embryonic m.e.p.s (Fig. 2*b*). However, embryonic and adult m.e.p.s injected into the same cell caused m.e.p.s of nearly the same amplitude (Fig. 2*b*). Therefore, the observed decrease in m.e.p.p. amplitudes during the first weeks after birth is probably not due to the appearance of the adult-type AChR. Preliminary modelling studies indicate that the reduction in m.e.p.p. amplitude is mostly due to a threefold increase in fibre length and diameter during the first three postnatal weeks.

To compare the relative efficiency of embryonic and adult-type AChRs in triggering action potentials, $1\text{--}9 \text{ nA}$ embryonic and adult m.e.p.s were injected into four 1–2-day-old fibres in which synaptic transmission had been blocked by $1 \mu\text{g ml}^{-1}$ α -bungarotoxin (α -BGT). Three of these cells fired action potentials and contracted vigorously in response to all ten injections of embryonic m.e.p.s ($4\text{--}5 \text{ nA}$). In these cells, adult m.e.p.s (9 nA) never elicited action potentials or muscle contractions. The fourth cell fired action potentials and contracted in response to even smaller (2 nA) embryonic m.e.p.s; adult m.e.p.s with amplitudes of 8 nA (well outside the observed $2.5\text{--}5.2 \text{ nA}$ range of spontaneous m.e.p.c. amplitudes) were required to excite this cell.

The long duration of embryonic m.e.p.s (due to the embryonic form of the AChR) is therefore essential for the appearance

of spontaneous contractions. However, the disappearance of spontaneous contractions does not depend on the switch to the adult type of the receptor because muscle fibre growth eventually overrides the ability of embryonic m.e.p.s to generate action potentials. A clue as to why receptors do switch to the adult type comes from the recently-described slow-channel syndrome disorder⁸. Affected patients have m.e.p.s that are $2\text{--}3$ times longer than normal, probably due to abnormally long AChR apparent mean channel open times. The myopathic changes observed have been linked to an excessive influx of Ca^{2+} through the AChR^{8–10}. The switch to the adult form of the receptor might be dictated, in part, by the need to prevent the harmful consequences of excessive Ca^{2+} accumulation in the muscle cell.

Why do muscle cells express the embryonic form of the AChR? We have shown here that the embryonic-type AChR contributes substantially to spontaneous muscle activity in rat soleus and that the adult form of the receptor could not play this role. Muscle activity promotes the differentiation of the neuromuscular junction and is essential for ontogenetic death of motor neurons, for the elimination of early multiple innervation, and for the initiation and maintenance of the differentiated state of muscle^{11–13}. Muscle activity is commonly thought to be due to evoked transmitter release from the nerve. However, it has recently been shown that activity due to spontaneous transmitter release promotes the differentiation of cultured *Xenopus* myocytes¹⁴. Our experiments in rat soleus suggest that the embryonic form of the AChR plays an important part in guaranteeing spontaneous contractile activity in developing muscles.

This paper is dedicated to the memory of the late Dr Stephen Schuetze. We thank Drs S. DeRiemer, D. Kelley, P. Middleton, L. Role, R. Pollock and C. Stevens for helpful comments and critical reading of the manuscript. This work was supported by grants to SMS from the NIH and the Muscular Dystrophy Association.

Received 13 April; accepted 8 July 1988.

- Maniatis, T., Fritsch, E. F., Lauer, J. & Lawn, R. M. *A. Rev. Genet.* **14**, 145-178 (1980).
- Cooper, A. T. & Ordahl, C. P. *J. biol. Chem.* **260**, 11140-11148 (1985).
- Sakmann, B. and Brenner, H. R. *Nature* **276**, 401-402 (1978).
- Fischbach, G. D. & Schuetze, S. *J. Physiol., Lond.* **303**, 125-137 (1980).
- Siegelbaum, S. A., Trautmann, A. & Koenig, J. *Dev. Biol.* **104**, 366-379 (1980).
- Vicini, S. & Schuetze, S. *J. Neurosci.* **5**, 2212-2224 (1985).
- Mishina, M. *et al. Nature* **321**, 406-411 (1986).
- Engel, A. G., Lambert, E. H. & Mulder, D. M. *Ann. Neurol.* **11**, 553-569 (1982).
- Leonard, J. P. & Salpeter, M. M. *J. Cell. Biol.* **82**, 811-819 (1979).
- Salpeter, M. M., Leonard, J. P. & Kasprzak, H. *Neurosci. Comment.* **1**, 73-83 (1982).
- Rubin, L. L., Schuetze, S. M., Weill, C. L. & Fischbach, G. D. *Nature* **283**, 264-267 (1980).
- Pittmann, R. & Oppenheim, R. W. *J. comp. Neurol.* **187**, 425-444 (1979).
- Lømo, T., Massoulié, J. & Vigny, M. *J. Neurosci.* **5**, 1180-1187 (1985).
- Kidokoro, Y. & Saito, M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 1978-1982 (1988).
- Land, B. R., Salpeter, E. E. & Salpeter, M. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3736-3740 (1980).

Drosophila nurse cells produce a posterior signal required for embryonic segmentation and polarity

Klaus Sander* & Ruth Lehmann†‡

* Biol. Institut I (Zoologie), Albertstr. 21, 76 Freiburg, FRG

† Max Planck Institut für Entwicklungsbiologie III, Spemannstr. 35, 74 Tübingen, FRG; and Medical Research Council, Laboratory of Molecular Biology, Cambridge CB2 2QH, UK

The segmental pattern of insect embryos depends on influences from morphogenetic centres near each of the egg poles¹. In *Drosophila*, maternal effect mutations are known that impair the normal function of each centre². Injection of wild-type cytoplasm into mutant eggs has revealed that morphogenetic signals localized at the anterior and posterior pole of eggs can be transplanted^{3,4}. We show here that these activities can also be detected during oogenesis. Posterior activity can be recovered at an early stage (stage 10, ref. 5) from the oocyte-nurse cell complex, but anterior activity can only be detected in the mature oocytes (stage 14). We conclude that the *bicoid*-dependent anterior signal, although produced by the nurse cells, does not become active before it is localized to the anterior egg pole, whereas posterior activity can be detected in the nurse cells before, and therefore independently of, its localization to the posterior egg pole.

Drosophila oogenesis starts with a germ cell which divides four times to yield a cluster of 16 cells connected by cytoplasmic bridges with the future oocyte in the most posterior position. The fifteen anterior cells form the nurse cell cluster, become polyploid and grow in volume until stage 10 (refs 5, 6). Subsequently, the nurse cells inject their cytoplasm in the oocyte⁷ and are completely absorbed by stage 14 when the egg becomes covered by the vitelline membrane and the outer chorion. Among the products of the nurse cells we would expect RNA transcripts required for the establishment of the segmental pattern in the embryo².

In the mature egg, centres of morphogenetic activity can be detected at the anterior and posterior pole by transplantation^{3,4,8}. Lack of the anterior centre leads to the loss of head, gnathal and thoracic structures, whereas lack of the posterior centre

Table 1 Stage-10 germ-line cells can rescue *oskar* embryos

Donor cytoplasm	Recipients		
	No. of cuticles analysed	Embryos rescued (%)	No. of abdominal segments (mean)
Stage 10 nurse cells	28	66	4.7
Stage 10 oocyte	18	67	1.8
Controls			
Posterior pole plasm	31	96	5.4
Uninjected	53	0	0

Cytoplasm from wild-type nurse cells, stage-10 oocytes, or the posterior pole of wild-type embryos, before pole cell formation, was injected into the prospective abdominal region (20-50% egg length) of early cleavage embryos derived from female homozygous mutant for the allele *osk*¹⁶⁶. The formation of abdominal denticle bands was used as a criterion for phenotypic rescue. Wild-type females fed on yeast for several days were dissected in Robb's saline¹⁸. Individual ovarioles or follicles were slightly compressed in a chamber made of two glue-covered cover-slips held apart by strips of adhesive tape. One side of the chamber was open for access and covered by Voltalet 10S oil to prevent evaporation of saline. Nurse cell cytoplasm (and possibly some nuclear material) was slowly taken up from the nurse cell cluster. Oocyte cytoplasm was taken from the entire stage-10 oocyte. Cytoplasm was collected from several nurse cell clusters or oocytes and injected into a series of mutant recipients. Preparation and injection of the recipient embryos were carried out as previously described³. The volume injected corresponded to about 5% or less of the recipient embryos.

causes the loss of abdominal segments⁸. These activities depend on the normal function of a small number of maternal effect genes². Embryos from females mutant for *bicoid* (*bcd*), the key gene for anterior development, develop no head and thorax: the abdominal region is extended and a second telson forms at the anterior (ref. 4, Fig. 1d). Embryos from females mutant for *oskar* (*osk*), one of the posterior group of genes, develop no abdomen, but head, thorax and telson are formed (ref. 3, Fig. 1b). The mutant phenotypes of *bcd* and *osk* embryos can be rescued by microinjection of wild-type egg cytoplasm^{3,4}. Rescuing activities for both phenotypes are strictly localized in the early embryo: the mutant *bcd* phenotype is rescued only by anterior cytoplasm and the abdominal *osk* phenotype is rescued only by posterior pole-plasm. Both groups of genes are transcribed in the female germ-line and their respective activities can be detected by transplantation as early as the unfertilized egg^{3,4}.

To detect anterior and posterior activity during oogenesis we transplanted cytoplasm from stage-10 nurse cells and oocytes into the abdominal region of mutant *osk* embryos and into the anterior region of mutant *bcd* embryos. We found that the abdominal phenotype of *osk* can be rescued with cytoplasm from stage-10 nurse cells and oocytes. Rescued embryos form up to the normal number of abdominal segments (Fig. 1c) and five to ten embryos can be rescued with cytoplasm from a single nurse cell cluster. Less activity is recovered from the oocyte at stage 10 (Table 1).

In contrast, no rescuing activity for anterior development can be detected at early stages of oogenesis (stage 10-13), the first stage at which we can detect activity being stage 14. Oocyte cytoplasm taken from the anterior pole of a mature oocyte (stage 14) caused 22% of injected *bcd* embryos to develop at least one thoracic segment and the anterior duplication of the telson was suppressed (number of embryos, *n* = 93). Thus wild-type follicles do not contain transplantable anterior activity before stage 14 and at this late stage of oogenesis the activity at the anterior pole is low compared to the activity at the anterior pole after egg deposition⁴.

Injection of stage-10 nurse cell cytoplasm does change the phenotype of *bcd* embryos dramatically. However, this effect is

‡ Present address: Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA.

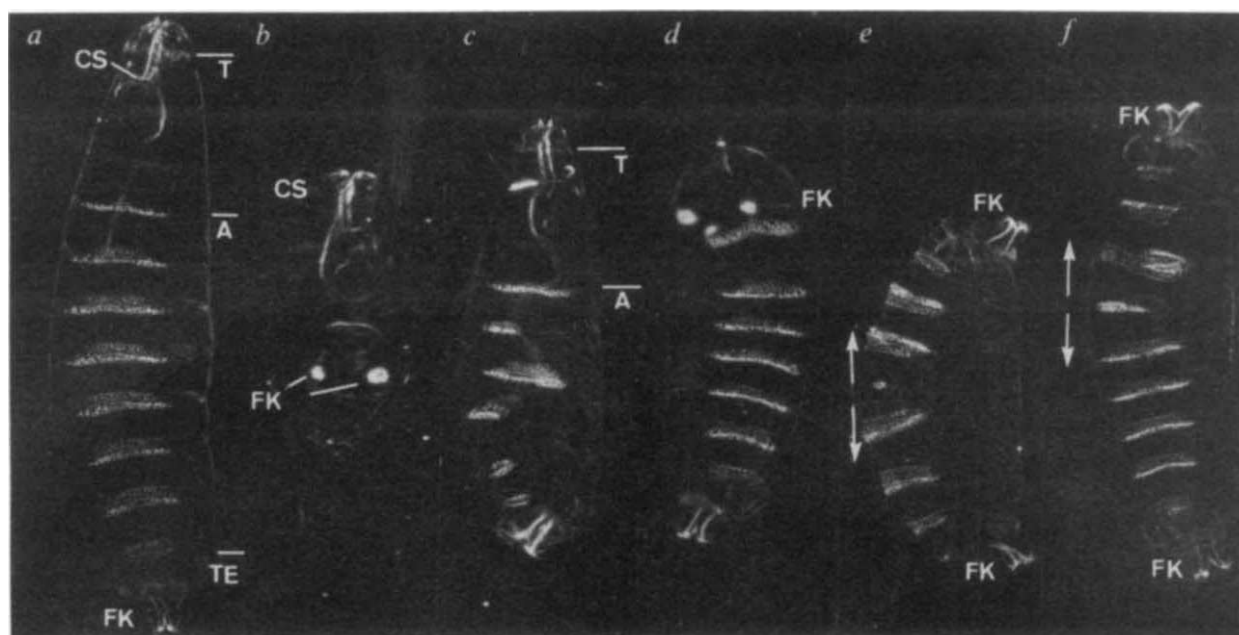


Fig. 1 Cuticle preparations of control and experimental embryos. *a*, Wild-type first instar larvae (see ref. 19 for description of wild-type pattern). *b*, Embryo derived from female homozygous mutant for *osk*¹⁶⁶. Head, thorax and telson formed normally but no abdomen developed. *c*, Embryo derived from *osk* female injected with nurse cell cytoplasm. This embryo developed six abdominal segments and segmentation is restored to different extents on the left and right side. *d*, Embryo derived from female heterozygous mutant for *bcd*^{E1} and *Df(3R) LIN* (ref. 4). Head structures and thorax are missing; instead, a telson is formed anteriorly. *e*, Embryo derived from *bcd* female injected with nurse cell cytoplasm into the anterior. This embryo developed two sets of abdominal segments in perfect mirror image (symmetrical double abdomen). The mirror plane is within the fifth abdominal denticle band. *f*, Embryo derived from *bcd* female and injected as embryo in *e*. This embryo developed an asymmetric double abdomen. Segmental polarity is reversed within a denticle band composed of the third abdominal segment of normal orientation and the fifth abdominal segment of reversed orientation.

Methods. Injections and cuticle preparations as described (ref. 3, Tables 1 and 2). Anterior top in all figs; all embryos are shown in frontal view on ventral side with exception of *e* which is a side view, ventral at the left. CS, cephalo-pharyngeal skeleton; T, thoracic region; A, abdominal region; TE, telson, posterior non-segmented region with Filzkörper (FK) which line the posterior opening of the tracheae. In *bcd* embryos anteriorly, and in *osk* embryos posteriorly, the Filzkörper are not normally formed because the eighth abdominal segment is lacking.

Arrows in *e* and *f* indicate segmental polarity and point anterior-posteriorly.

probably caused by posterior activity present in the nurse cells because posterior pole plasm transplanted into the anterior of *bcd* embryos produced the same effect². Almost all *bcd* embryos injected with nurse cell cytoplasm develop a second set of abdominal segments (Fig. 1*e, f*) and 65% of the embryos form two posterior abdomens in perfect mirror image symmetry (Fig. 1*e*, Table 2) reminiscent of *bicaudal* mutants⁹⁻¹¹. Our trans-

plantations into *bicoid* and *oskar* embryos both show that nurse cell cytoplasm contains activity for the induction of posterior development but not for the induction of anterior development.

The transplantation experiments indicate that the mechanisms by which the anterior (*bcd* dependent) and posterior (*osk* dependent) factors acquire activity in the oocyte are different. Anterior activity cannot be recovered from the nurse cell cluster. This is surprising because *bcd* RNA is expressed in all nurse cells as early as stage 10^{12,13}. An inhibitory effect of posterior activity present in the nurse cells on the morphogenetic function of *bcd* was excluded by experiments using cytoplasm from mutant *nanos* (*nos*, ref. 2) nurse cells which lack posterior activity (R.L., in preparation). To make the assay sensitive for anterior as well as residual posterior activity *nos* nurse cell cytoplasm was injected in embryos derived from *bcd nos* double mutant females. Uninjected *bcd nos* double mutant embryos develop two telsons in mirror image and lack segmentation (see *bcd osk* in ref. 2). After injection double mutant embryos developed the same phenotype as the uninjected control embryos. Neither anterior segmentation (head, thorax) nor posterior segmentation (abdomen) was restored (*n* = 25). Therefore *bcd* RNA seems to be inaccessible or inactive in the nurse cells. At the anterior pole of the oocyte we can detect activity once the oocyte is mature (stage 14) only, although *bcd* RNA is found at the anterior of the developing oocyte as early as stage 8 (ref. 13). Thus *bcd* RNA may be activated during the last stage of oogenesis. Alternatively, the *bcd* protein may be the rescuing principle, but as the *bcd* protein has not been observed before egg deposition¹⁴ we would have to assume that the protein is highly unstable during oogenesis or that small amounts of protein in the oocyte (just enough for minimal rescue) have escaped detection.

Table 2 Stage-10 nurse cells can cause double abdomen formation in *bicoid* embryos

Donors	Maternal genotype	Recipients	
		No. of cuticles analysed	% Double abdomens (% symmetrical)
Wild-type nurse cells	<i>bicoid</i>	31	90 (71)
Controls			
No injection	<i>bicoid</i>	150	4 (75)
Injection of Robb's saline	<i>bicoid</i>	71	2 (100)

Nurse cell cytoplasm from wild-type ovaries was injected into the anterior tip of *bcd*. Our definition of double-abdomen is the production of abdominal segments at the anterior in reverse polarity to the normal abdomen. In symmetric double abdomens the same number of segments is formed on both sides of the plane of mirror image (cf. Fig. 1*e*). Up to three double abdomens were obtained from the cytoplasm of a single nurse cell group. Untreated *bcd* embryos rarely produce double abdomens (see controls). The percentage of symmetric double abdomens among all double abdomens is given in parentheses. Genotype of females used was *Df(3R) LIN/bcd*^{E1}. The volume of Robb's saline¹⁸ injected corresponded to at least 5% egg volume. Preparation of ovarioles and follicles was as described in Table 1.

Unlike the anterior activity, posterior activity is initially distributed throughout the stage-10 oocyte-nurse cell complex and becomes localized to the posterior pole at the end of oogenesis. In the nurse cells, posterior activity is abundant and several *osk* embryos can be rescued with the cytoplasm of a single nurse-cell cluster. This surplus activity might enable posterior factors to reach the posterior pole and be maintained at high concentrations there as it decays elsewhere in the egg. At the posterior pole the morphologically specialized posterior pole plasm¹⁵ may selectively stabilize the posterior activity. Various genetic and developmental experiments indicate a correlation between the existence of a specialized pole plasm with polar granules and abdominal segmentation and polarity. Embryos from females mutant for *osk* or four other posterior group genes (*vasa*, *staufer*, *valois*, *tudor*, refs 16, 17) do not have a specialized pole plasm. Such embryos lack posterior activity at the posterior pole and therefore do not form abdominal segments (3, R.L., unpublished data).

We would like to thank Peter Lawrence and Maria Leptin for stimulating discussions of the manuscript. K.S. wishes to thank

Christiane Nüsslein-Volhard and her group for their hospitality during his sabbatical, Hans Georg Frohnhöfer for help and the Deutsche Forschungsgemeinschaft for financial support.

Received 12 May; accepted 27 July 1988.

1. Sander, K. *Adv. insect. Physiol.* **12**, 125-258 (1976).
2. Nüsslein-Volhard, C., Frohnhöfer, H. G. & Lehmann, R. *Science* **238**, 1675-1681 (1987).
3. Lehmann, R. & Nüsslein-Volhard, C. *Cell* **47**, 141-152 (1986).
4. Frohnhöfer, H. G. & Nüsslein-Volhard, C. *Nature* **324**, 120-125 (1986).
5. King, R. C. *Ovarian Development in Drosophila melanogaster* (Academic, New York, 1970).
6. Bohrmann, J. & Sander, K. *Wilhelm Roux Arch. dev. Biol.* **196**, 279-285 (1987).
7. Gutzeit, H. O. & Koppa, R. *J. Embryol. exp. Morph.* **67**, 101-111 (1982).
8. Frohnhöfer, H. G., Lehmann, R. & Nüsslein-Volhard, C. *J. Embryol. exp. Morph.* **97**, 169-179 (1986).
9. Bull, A. J. *Expl. Zool.* **161**, 221-242 (1966).
10. Nüsslein-Volhard, C. *Wilhelm Roux Arch. dev. Biol.* **183**, 249-268 (1977).
11. Mohler, J. & Wieschaus, E. *Genetics* **112**, 803-822 (1985).
12. Frigerio, G., Burri, M., Bopp, D., Baumgartner, S. & Noll, M. *Cell* **47**, 735-746 (1986).
13. Berleth, T. *et al.* *EMBO J.* **7**, 1749-1756 (1988).
14. Driever, W. & Nüsslein-Volhard, C. *Cell* **54**, 83-93 (1988).
15. Mahowald, A. P. *J. Expl. Zool.* **151**, 201-215 (1962).
16. Schüpbach, T. & Wieschaus, E. *Wilhelm Roux Arch. dev. Biol.* **191**, 302-317 (1986).
17. Boswell, R. E. & Mahowald, A. P. *Cell* **43**, 97-104 (1985).
18. Robb, J. A. *J. Cell. Biol.* **41**, 876-885 (1969).
19. Lohs-Schardin, M., Cremer, C. & Nüsslein-Volhard, C. *Dev. Biol.* **73**, 239-255 (1979).

Ciliary neurotrophic factor induces type-2 astrocyte differentiation in culture

Simon M. Hughes*†, Laura E. Lillien*,
Martin C. Raff*, Hermann Rohrer‡
& Michael Sendtner‡

* MRC Developmental Neurobiology Programme,
Department of Biology, Medawar Building,
University College London, London WC1E 6BT, UK

‡ Department of Neurochemistry,
Max-Planck Institute for Psychiatry, Am Klopferspitz 18a,
D-8033 Martinsried, FRG

We have been studying a population of bipotential glial progenitor cells in the perinatal rat optic nerve and brain in an attempt to understand how cells choose between alternative fates in the developing mammalian central nervous system (CNS). This cell population gives rise initially to oligodendrocytes and then to type-2 astrocytes¹, both of which apparently collaborate in sheathing axons in the CNS^{2,3}. *In vitro* studies suggest that oligodendrocyte differentiation is the constitutive pathway of development for the oligodendrocyte-type-2-astrocyte (O-2A) progenitor cell^{4,5}, whereas type-2 astrocyte differentiation depends on a specific inducing protein⁶. This protein is present in the developing optic nerve when type-2 astrocytes are differentiating and can induce O-2A progenitor cells *in vitro* to express glial fibrillary acidic protein (GFAP)⁶, a marker of astrocyte differentiation⁷. Here we show that the type-2-astrocyte-inducing protein is similar or identical to ciliary neurotrophic factor (CNTF)^{8,9}, which promotes the survival of some types of peripheral neurons *in vitro*⁸, including ciliary ganglion neurons^{8,10}. This suggests that CNTF, in addition to its effect on neurons, may be responsible for triggering type-2 astrocyte differentiation in the developing CNS.

CNTF was purified from rat sciatic nerve using a modification of the method used previously⁹. When it was added to cultures of postnatal-day-0 (P0) rat optic nerve cells it induced GFAP expression in about 20% of O-2A progenitor cells (Fig. 1a). The same result was obtained with an extract of 3-4-week-old rat optic nerve (Fig. 1b), as described previously⁶. CNTF and optic nerve extract had the same effect when added simul-

Table 1 The GFAP-inducing effect of CNTF is transient and decreases with increasing age of optic nerve

Age of rats	Days <i>in vitro</i>	% O-2A progenitor cells expressing GFAP		
		With CNTF (1-4 ng ml ⁻¹)	With optic nerve extract (2 µg ml ⁻¹)	With neither
E18	1	35 ± 2	30 ± 4	0
P0	1	21 ± 5	19 ± 4	0
P7	1	9 ± 5	8 ± 4	0.2 ± 0.5
P0	1	18 ± 7	18 ± 5	0
P0	3	0	0	0

Optic nerve cells were prepared, cultured, and assayed as described in Fig. 1. Results are expressed as mean ± standard deviation of at least three cultures.

taneously as when either was added alone (Fig. 1b). Two other neurotrophic factors, nerve growth factor^{11,12} and brain-derived growth factor¹³, had no effect on GFAP expression in this assay (not shown).

CNTF and the type-2-astrocyte-inducing protein share several properties. They are both acidic (ref. 9 and unpublished observations), with a relative molecular mass (*M_r*) of about 25,000 (25K) (refs 6 and 9), and present at relatively high concentrations in sciatic nerve^{6,14} and kidney (ref. 9 and unpublished observations), compared with liver, retina or brain (refs 6, 14 and unpublished observations). If the type-2-astrocyte-inducing activity in optic nerve extracts is CNTF, then these extracts should support ciliary ganglion neuron survival *in vitro*, which they did (Fig. 1d). The responses of both ciliary ganglion neurons and O-2A progenitor cells were elicited within a similar range of concentrations (Fig. 1); the specific activity of optic nerve extract was about 1,300-fold less than that of CNTF in both assays.

When tested on optic nerve cells CNTF had many of the functional properties previously described for optic nerve extract⁶. First, CNTF had no appreciable effect on GFAP expression in type-1 astrocytes or their precursors in cultures of embryonic day 17 (E17) optic nerve cells, even though it induced many of the O-2A progenitor cells in such cultures to express GFAP (not shown). Second, unlike fetal calf serum (FCS), which is also an inducer of type-2 astrocyte differentiation *in vitro*¹, CNTF induced only a proportion of O-2A progenitor cells in optic nerve to express GFAP, and this proportion decreased with increasing age from E18 to P7 (Table 1). Third,

† Present address: Department of Pharmacology, Stanford University Medical School, Stanford, California 94305, USA.

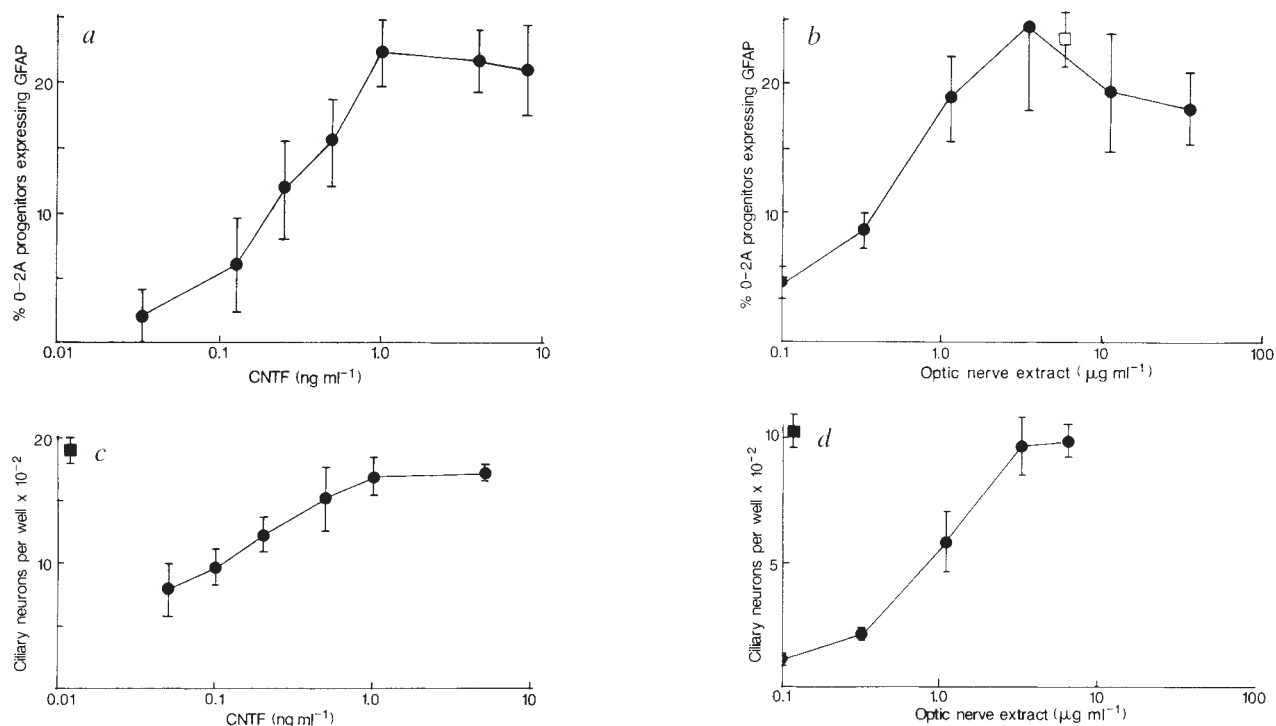


Fig. 1 The effects *in vitro* of sciatic nerve CNTF and optic nerve extract on GFAP expression in O-2A progenitor cells (*a* and *b*) and on survival of ciliary ganglion neurons (*c* and *d*).

Methods. Optic nerve cells from newborn (P0) Sprague-Dawley (S-D) rats were prepared as previously described¹⁶ except that the dissociation was by 10 passages through a Pasteur pipette and the cells were washed in Dulbecco's modified Eagles' medium (DMEM) containing a variety of additives (modified from Bottenstein and Sato²⁴ as previously described⁴) and 0.5% fetal calf serum. The cells were cultured in the same medium on poly-D-lysine-coated coverslips (about 8,000 cells per coverslip) with CNTF, or optic nerve extract or both (□, 5 ng ml⁻¹ of CNTF and 8 μg ml⁻¹ of extract) for 24 h. After fixation in 4% paraformaldehyde in phosphate-buffered saline for 5 min at room temperature, the cells were double-labelled with the A2B5 monoclonal antibody²⁵ followed by goat anti-mouse immunoglobulin G (IgG) coupled to rhodamine (Cappel, 1:100), and then (after postfixation in acid-alcohol) rabbit anti-GFAP serum²⁶ (1:1000) followed by sheep anti-rabbit IgG coupled to fluorescein (Wellcome, 1:100), as previously described⁶. The cells were then mounted and examined in a fluorescence microscope to determine the proportion of O-2A progenitor cells (recognized by their process-bearing morphology and staining with the A2B5 antibody^{1,27}) that expressed GFAP as described previously⁶. Ciliary ganglion neurons were prepared from 8-day-old chick embryos and cultured at a density of 1–2 × 10³ cells well⁻¹ in multiwell dishes (Costar, 16 mm) in F14 medium with 10% horse serum, as previously described²⁸, except that the culture dishes were coated with laminin (BRL, 2 μg in 500 μl per well) instead of heart-conditioned medium. The number of neurons initially plated per well was determined by counting the neurons 2 h after plating (■). After 24 h the number of surviving neurons was counted. Each point is the mean of at least 3 determinations (except for the two highest concentrations in Fig. 3b which were done in duplicate) and the bars represent the standard errors. Extracts of optic nerves from 3–4-week-old S-D rats were prepared as previously described⁶. CNTF was prepared from adult Wistar rats by a modification of the method of Manthorpe and co-workers^{8,9}, using DEAE ion-exchange chromatography and preparative SDS-PAGE. The purified CNTF gave a single silver-stained band when analysed by SDS-PAGE (see Fig. 2) or isoelectric focussing (not shown); partial amino acid sequence analysis of fragments of the purified protein has shown that the preparation contains a single protein. Both CNTF and optic nerve extract lost biological activity with freezing and thawing, which probably accounts for the small differences in the dose-response curves between *a* and *c* and between *b* and *d*.

and again unlike FCS, CNTF caused only transient GFAP expression in O-2A progenitor cells, even when added daily: the response was maximal by 1 day and gone by 3 days (Table 1). This suggests that CNTF can initiate type-2 astrocyte differentiation but cannot on its own induce O-2A progenitor cells to complete the process. Both the age-dependence and transient nature of the response are discussed elsewhere (ref. 6 and L.E.L. and M.C.R., in preparation).

The type-2-astrocyte-inducing activity in optic nerve extract co-migrated with purified CNTF on SDS-polyacrylamide gel electrophoresis (SDS-PAGE) (Fig. 2). When an extract of 3–4-week-old rat optic nerve was submitted to the same purification procedure used to obtain sequence-pure CNTF from rat sciatic nerve extract, several polypeptides of $M_r \approx 25K$ were obtained. This semi-purified preparation from optic nerve had type-2-astrocyte-inducing activity and promoted ciliary ganglion neuron survival (Fig. 3), and the specific activity was about 10-fold less than that of CNTF purified from sciatic nerve (compare Figs 1 and 3). Thus a protein with very similar properties to CNTF is present in optic nerve, but it could not be

purified by this procedure, at least partly because it is apparently present at much lower concentrations in optic nerve than in sciatic nerve: in both type-2 astrocyte induction and neuronal survival assays, crude extracts of sciatic nerve had at least a 10-fold higher specific activity than crude extracts of optic nerve (data not shown).

We showed previously that extracts of optic nerve from 3–4-week-old rats have 20- to 100-fold higher type-2-astrocyte-inducing activity than extracts of optic nerve from 1-week-old rats, when normalized for total protein concentration⁶. When tested in the ciliary ganglion neuron survival assay, extracts of 3–4-week-old nerves had a half-maximal effect at around 1 μg ml⁻¹ of total protein whereas an extract of 4–5-day-old nerves had a half-maximal effect at around 25 μg ml⁻¹.

Taken together, our results indicate that the type-2-astrocyte-inducing protein in optic nerve is CNTF or a closely related protein. Several lines of evidence suggest that this protein is responsible for timing the onset of type-2 astrocyte differentiation. (1) It greatly increases in the optic nerve between the first and third postnatal weeks⁶, when type-2 astrocytes develop *in*

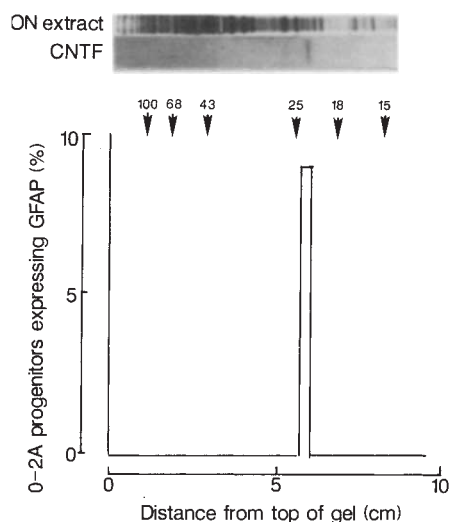


Fig. 2 The type-2-astrocyte-inducing activity in optic nerve extract co-migrates with CNTF on SDS-PAGE

Methods. CNTF and an extract from 3-4-week-old rat optic nerve were fractionated by electrophoresis on a 15% SDS polyacrylamide gel under non-reducing conditions²⁹, along with standard relative molecular mass markers, as previously described⁶. One track containing crude optic nerve extract (7 μ g of protein) and another containing 20 ng of pure CNTF were silver-stained³⁰ and are shown at the top of the figure. Another track from the same gel, containing crude optic nerve extract (100 μ g of protein) was cut into 2.5 mm slices and the protein in each slice was eluted in 0.5 ml DMEM containing 0.5 mg ml⁻¹ bovine serum albumin³¹, as previously described⁶; the protein eluted from each slice was tested (at a final dilution of 1:10) on cultures of PO optic nerve cells for GFAP-inducing activity as described in Fig. 1. The arrowheads indicate the positions of standard markers (BRL), while the numbers over the arrowheads indicate the M_r ($\times 10^{-3}$) of the markers.

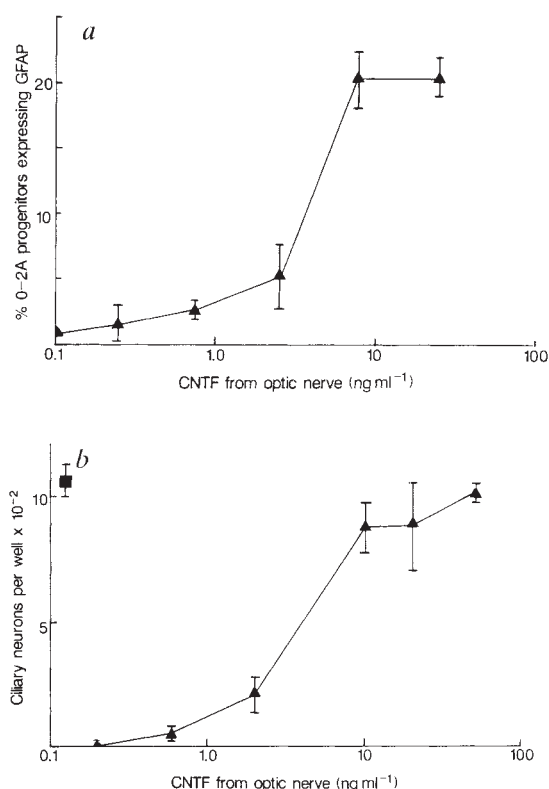


Fig. 3 The effects *in vitro* of CNTF prepared from optic nerve on GFAP expression in O-2A progenitor cells (a) and on survival of ciliary ganglion neurons (b).

Methods. CNTF was prepared from the optic nerves of 3-4-week-old S-D rats and assayed as described in Fig. 1. Optic nerve cells were prepared as described in Fig. 1 or as previously described⁴.

*vivo*¹⁶ (2) It acts on O-2A progenitor cells isolated from E17-18 optic nerve⁶, which is more than 10 days before these cells normally develop into type-2 astrocytes¹⁶. This suggests that it is the increase in the CNTF-like protein rather than the onset of progenitor cell responsiveness to the protein that is responsible for timing type-2 astrocyte differentiation in the developing optic nerve. (3) A similar molecule with type-2-astrocyte-inducing activity is released by cultures of embryonic rat brain starting at the time type-2 astrocytes begin to develop; extracts of such cultures have much more of this activity and also promote ciliary ganglion neuron survival, while extracts prepared from cultures before type-2 astrocytes have appeared have at least 50-fold less of both activities¹⁷.

Studies on the development of optic nerve glial cells in culture have suggested that type-1 astrocytes, the first glial cells to develop in the nerve¹⁶, play a crucial role in timing the start of oligodendrocyte differentiation^{4,18} by secreting platelet-derived growth factor (PDGF)^{15,19,20}. The PDGF stimulates O-2A progenitor cells to proliferate^{15,18} and drives an intrinsic clock in the progenitor cells that determines when they will differentiate into oligodendrocytes²⁰. There is evidence that type-1 astrocytes in culture can also synthesize CNTF or a closely related molecule²¹, which can initiate differentiation along the type-2 astrocyte pathway¹⁷. Thus, by secreting PDGF and CNTF, type-1 astrocytes may regulate the timing of both oligodendrocyte and type-2 astrocyte development. The ability to reconstitute this timing *in vitro*^{17,22} should facilitate the task of determining how the synthesis (or release) of these proteins is regulated.

CNTF was originally identified¹⁰ and isolated^{8,9} as a factor that promotes the survival of postmitotic peripheral neurons *in vitro*. It has recently been shown that CNTF also affects the

proliferation²³ and differentiation (H.R. and U. Ernsberger, submitted) of E7 chick sympathetic neurons *in vitro*. These findings, together with the results presented here, suggest a role for CNTF in the control of both neuronal and glial differentiation.

We thank H. Thoenen and Y.-A. Barde for helpful discussions, A. Mudge for helpful comments on the manuscript, and Eva Schmitz for excellent technical assistance. S.M.H. was supported by a grant from the National Fund for Research into Crippling Diseases and L.E.L. by a Fellowship from the National Multiple Sclerosis Society of the United States.

Received 20 May; accepted 13 July 1988.

1. Raff, M. C., Miller, R. & Noble, M. *Nature* **303**, 390-396 (1983).
2. French-Constant, C. & Raff, M. C. *Nature* **323**, 335-338 (1986).
3. Miller, R. H., French-Constant, C. & Raff, M. C. *A. Rev. Neurosci.* (in the press).
4. Raff, M. C., Abney, E. R. & Fok-Seang, J. *Cell* **42**, 61-69 (1985).
5. Temple, S. & Raff, M. C. *Nature* **313**, 223-225 (1985).
6. Hughes, S. M. & Raff, M. C. *Development* **101**, 157-167 (1987).
7. Bignami, A., Eng, L. F., Dahl, D. & Uyeda, C. T. *Brain Res.* **43**, 429-435 (1972).
8. Barbin, G., Manthorpe, M. & Varon, S. *J. Neurochem.* **43**, 1468-1478 (1984).
9. Manthorpe, M., Skaper, S. D., Williams, L. R. & Varon, S. *Brain Res.* **367**, 282-286 (1986).
10. Helfand, S., Smith, G. A. & Wessells, N. K. *Dev. Biol.* **50**, 541-547 (1976).
11. Greene, L. A. & Shooter, E. M. *A. Rev. Neurosci.* **3**, 353-402 (1980).
12. Levi-Montalcini, R. *Science* **237**, 1154-1162 (1987).
13. Barde, Y.-A., Edgar, D. & Thoenen, H. *EMBO J.* **1**, 549-553 (1982).
14. Rudge, J. S., Davis, G. E., Manthorpe, M. & Varon, S. *Dev. Brain Res.* **32**, 103-110 (1987).
15. Richardson, W. D., Pringle, N., Mosley, M., Westermark, B. & Dubois-Dalcq, M. *Cell* **53**, 309-319 (1988).
16. Miller, R. H., David, S., Patel, R., Abney, E. R. & Raff, M. C. *Dev. Biol.* **111**, 35-41 (1985).
17. Lillien, L. E., Sendtner, M., Rohrer, H., Hughes, S. M. & Raff, M. C. *Neuron* (in the press).
18. Noble, M. & Murray, K. *EMBO J.* **3**, 2243-2247 (1984).
19. Noble, M., Murray, K., Stroobant, P., Waterfield, M. & Riddle, P. *Nature* **333**, 559-561 (1988).
20. Raff, M. C., Lillien, L. E., Richardson, W., Burnes, J. & Noble, M. *Nature* **333**, 562-565 (1988).
21. Rudge, J. S., Manthorpe, M. & Varon, S. *Dev. Brain Res.* **19**, 161-172 (1985).
22. Williams, B. P., Abney, E. R. & Raff, M. C. *Dev. Biol.* **112**, 126-134 (1985).

23. Rohrer, H., Sendtner, M. & Ernsberger, U. *Biol. Chem. Hoppe Seyler* **368**, 1290-1291 (1987).
24. Bottenstein, J. E. & Sato, G. H. *Proc. natn. Acad. Sci. U.S.A.* **76**, 514-517 (1979).
25. Eisenbarth, G. S., Walsh, F. S. & Nirenberg, M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4913-4917 (1979).
26. Pruss, R. *Nature* **280**, 688-690 (1979).
27. Temple, S. & Raff, M. C. *Cell* **44**, 773-779 (1986).
28. Harper, G. P. *et al. Neuroscience* **8**, 375-387 (1983).
29. Laemmli, U. K. & Favre, M. *J. molec. Biol.* **80**, 577-599 (1974).
30. Oakley, B. R., Kirsch, D. R. & Morris, N. R. *Analyt. Biochem.* **105**, 361-363 (1980).
31. Lemke, G. E. & Brookes, J. P. *J. Neurosci.* **4**, 75-83 (1984).

Substance P regulates the vasodilator activity of calcitonin gene-related peptide

S. D. Brain & T. J. Williams

Section of Vascular Biology, MRC Clinical Research Centre, Watford Road, Harrow, Middlesex, HA1 3UJ, UK

The 37-amino-acid calcitonin gene-related peptide (CGRP) occurs as a result of alternative processing of mRNA from the calcitonin gene¹. The potency of CGRP as a vasodilator² and the occurrence of the peptide in nerves associated with blood vessels^{1,3} suggest an important role for CGRP in the regulation of blood flow. The finding that CGRP induces protracted vasodilatation when administered extra-vascularly^{2,4}, to mimic release from nerves, has led us to investigate how the vasodilator activity of CGRP is controlled *in vivo*. CGRP is often co-localized with substance P in C-fibre nerves^{5,6}. Here, we demonstrate that injection of CGRP with substance P into human skin converts the long-lasting vasodilatation induced by CGRP into a transient response. Experiments in animals reveal that the phenomenon is dependent on the action of proteases from mast cells stimulated by substance P. The results reveal a new regulatory interaction between two neuropeptides and provide evidence for an *in vivo* role for mast cell proteases.

Figure 1a shows the protracted response to intradermally injected CGRP (10 pmol in 0.05 ml) in the human forearm. Local erythema, clearly visible over the 4-hour experimental period, was associated with a prolonged increase in microvascular blood flow (Fig. 1b). Erythema responses gradually increased in diameter and, occasionally, long pseudopodia developed, probably because of movement of the peptide along cutaneous lymphatics². This implies that CGRP retains biological activity for long periods in cutaneous tissue fluid. We have also found that the activity of the peptide is stable upon incubation in blood^{4,7}.

We have suggested that endogenous CGRP could have a role in the regulation of blood flow^{2,4}. One criterion⁸ in assigning such a role is that mechanisms should exist to metabolize CGRP, so that its biological activity can be controlled. CGRP is often co-localized with substance P (11 amino acids) in nerves^{5,6,9}. Intradermally injected substance P induced an immediate wheal and surrounding flare which disappeared within 30 minutes, as previously observed¹⁰. However, the intradermal injection of CGRP together with substance P, used to simulate co-release from nerve terminals, converted the protracted vasodilator response of CGRP to a transient response in terms of both erythema (Fig. 1a) and cutaneous blood flow (Fig. 1b). Substance P releases mast cell histamine¹¹ but a dose of histamine, matched to induce a similar wheal and flare response⁴, did not influence the protracted vasodilatation when mixed with CGRP (Fig. 1b). Neurokinin A (NKA) and neurokinin B (NKB) have the same C-terminal amino-acid sequence as substance P, but lack the basic N-terminus which activates mast cells^{12,13}. NKA and NKB both caused acute local wheal formation (stimulated by the C-terminal sequence)¹², but neither agent affected the vasodilator activity of CGRP (Fig. 1). These results indicate that substance P attenuates the activity of CGRP by a mechanism

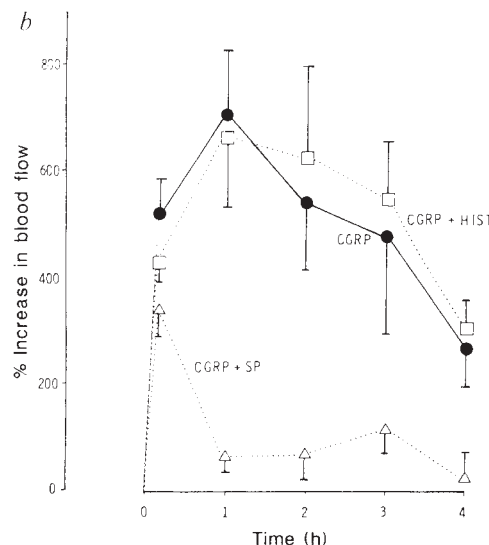
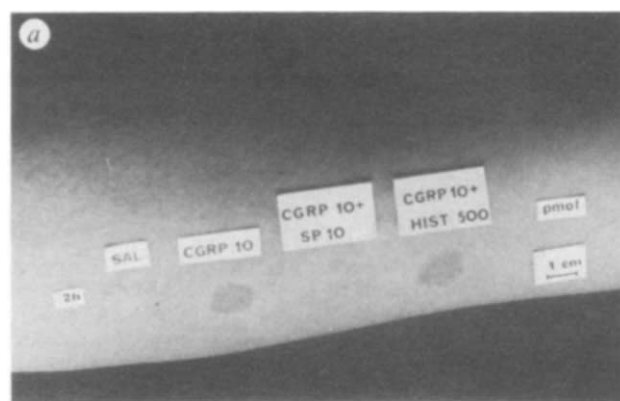


Fig. 1 Loss of CGRP-induced vasodilator activity at sites injected with substance P in human skin. **a**, The response of the human forearm 2 h after intradermal injections. Sites received: saline (SAL), CGRP (10 pmol) alone, CGRP (10 pmol) + substance P (SP, 10 pmol) and CGRP (10 pmol) + histamine (HIST, 500 pmol). Well defined erythema was observed at the sites which received CGRP alone and CGRP + HIST. In contrast, little erythema was apparent at the (CGRP + SP)-injected site. **b**, The effect of agents on skin blood flow. The ordinate shows the percentage increase in blood flow at injected sites compared with blood flow measured at those sites before injection. Blood flow remained stable (2% mean change) at uninjected sites throughout the experiment. At 10 min after injection, blood flow was raised at all CGRP-treated sites. At this time, local erythema was observed at CGRP-injected sites and local erythema plus the acute wheal and flare was observed at (CGRP + HIST)- and (CGRP + SP)-injected sites. From 1-4 h after injection blood flow remained high at sites which received CGRP (●) and (CGRP + HIST) (□) but not at sites which received (CGRP + SP) (△). Results are expressed as mean $\pm$ s.e.m. (standard error of the mean of $n=6$ individuals for 0-2 h and $n=5$ individuals for 3-4 h). In a second series of experiments, intradermal NKA (10 pmol), NKB (10 pmol) and SP (10 pmol) each caused an acute wheal (at 10 min: SAL, 18.2 ± 7.8 mm²; NKA, 83.4 ± 5.9 mm²; NKB, 61.2 ± 6.2 mm²; SP, 70.6 ± 7.7 mm², results are expressed as area, mean $\pm$ s.e.m., $n=4$). When injected with CGRP (10 pmol) only SP caused a loss of the prolonged vasodilator response (at 1 h: CGRP, $613 \pm 57\%$; CGRP + NKA, $597 \pm 104\%$; CGRP + NKB, $541 \pm 103\%$; CGRP + SP, $117 \pm 67\%$, results are expressed as % increased blood flow, mean $\pm$ s.e.m., $n=4$). **Methods.** Four sites on the volar surface of the forearm of six normal individuals (5 male and 1 female, 26-41 yr) were injected intradermally with 0.05 ml volumes¹⁰, according to a protocol approved by the local ethical committee. Human synthetic α -CGRP (Bachem) was used in all experiments. Local blood flow was measured at injected sites using a laser Doppler blood flow meter^{4,18} before injections and at 10 min, 1 h, 2 h, 3 h and 4 h, after intradermal injections.

Fig. 2 Loss of the vasodilator activity of CGRP when injected with substance P, C.48/80 or proteases into rat skin. *a*, Effect of substance P. Increased blood flow induced by CGRP (20 pmol per site, ●) was attenuated in a dose-related manner by the co-injection of substance P (1–100 pmol, △). Substance P alone (1–100 pmol, ▲) did not increase blood flow. Results are expressed as the mean $\pm$ s.e.m. for the number of rats indicated. Oedema formation measured at injected sites (μ l plasma per skin site) was as follows: PBS, 9.2 ± 0.4 μ l; CGRP (20 pmol), 10.8 ± 0.6 μ l; SP (1 pmol), 15.6 ± 2.4 μ l; SP (1 pmol) + CGRP, 26.6 ± 5.3 μ l; SP (10 pmol), 14.6 ± 1.1 μ l; SP (10 pmol) + CGRP, 34.0 ± 2.5 μ l; SP (100 pmol), 26.1 ± 2.4 μ l; SP (100 pmol) + CGRP, 58.2 ± 5.8 μ l. In initial experiments, where the effect of pretreatment with amine antagonists on oedema responses was investigated, local oedema induced by substance P (10 pmol) + CGRP (20 pmol) was inhibited by $57.8 \pm 7.3\%$ in pretreated animals, ($p < 0.05$, paired t -test $n = 4$). Oedema induced by compound 48/80 was completely inhibited in pretreated animals (see below). *b*, Effect of C.48/80. C.48/80 (10 μ g per site) abolished the increased blood flow induced by CGRP (20 pmol per site). This effect was prevented when soya bean trypsin inhibitor (SBTI, 50 nmol per site) was added to the injection mixture (stippled columns). Results are expressed as a mean $\pm$ s.e.m. from 6 rats. C.48/80, unlike substance P (Fig. 2*a*), did not induce oedema (PBS, 10.7 ± 1.8 μ l; C.48/80 (10 μ g), 15.2 ± 3.3 μ l; C.48/80 + CGRP, 16.8 ± 4.0 μ l). *c*, Effect of trypsin and chymotrypsin. CGRP (20 pmol per 0.1 ml) was incubated with trypsin (from bovine pancreas, 1 μ g per 0.1 ml) and chymotrypsin, (from bovine pancreas 1 μ g per 0.1 ml) in PBS for 15 min at 37 °C with and without SBTI (50 nmol per 0.1 ml). Incubations were terminated by cooling to 0 °C and adding SBTI to tubes which had not received SBTI before incubation. Samples were then injected into rat skin and blood flow measured at each site. Results are expressed as above and are the mean $\pm$ s.e.m. from 4 experiments. Trypsin (T) and chymotrypsin (Ch) abolished the vasodilator activity of CGRP and this was prevented in both cases, by the addition of SBTI before incubation.

Methods. The dorsal skin of Sprague Dawley rats (200–500 g) was shaved and depilated. Rats were pretreated intra-peritoneally with mepyramine (3 mg kg⁻¹) and methysergide (6 mg kg⁻¹) 15 min before intradermal injections. Intradermal injections (0.1 ml) of the agents under test in phosphate-buffered saline (PBS) were given at 2-min intervals according to a balanced site pattern with 2 replicates for each agent in each animal. Three blood flow readings (each for 15 s) were taken sequentially at each site, 30 min after intradermal injection. Thus, six readings provided the mean response to a particular treatment in each rat. Results are shown in the figure as % change in blood flow from that observed in PBS-injected sites. [¹²⁵I] human serum albumin (5 μ Ci) was administered intravenously before intradermal injections so that oedema formation could also be measured at injected skin sites⁸.

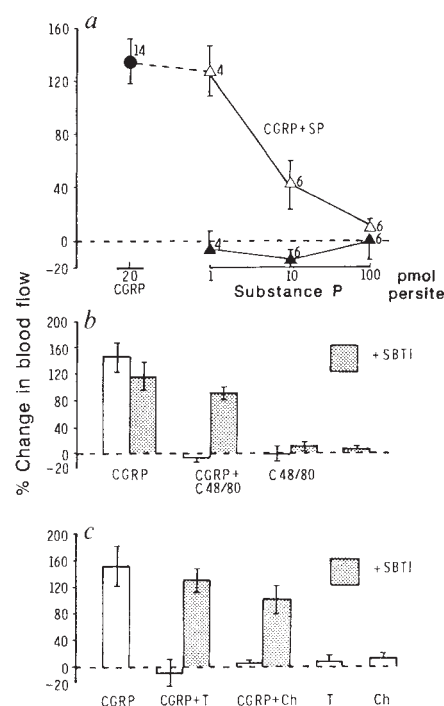
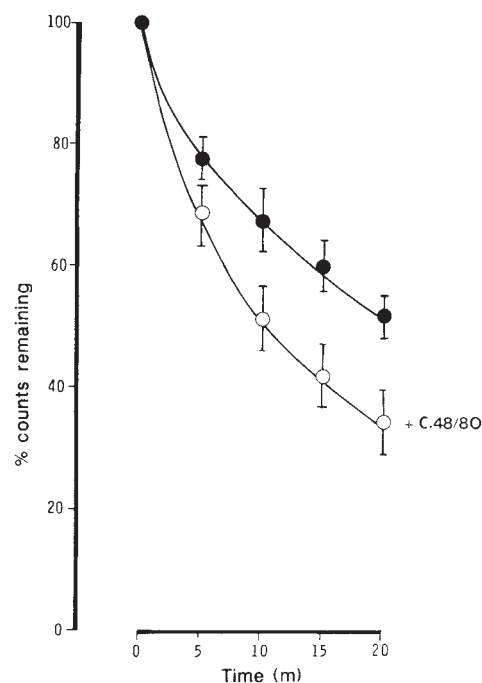


Fig. 3 Increased clearance of [¹²⁵I] CGRP from rat skin in the presence of C.48/80. Rats were pretreated with mepyramine and methysergide as for Fig. 2. (2-[¹²⁵I]-iodohistidyl)CGRP was made up in PBS containing the histamine H₂ antagonist cimetidine (100 ng per 0.1 ml) and the prostaglandin synthetase inhibitor indomethacin (100 ng per 0.1 ml). Cimetidine and indomethacin were used in these experiments to ensure that there was no contributing increased blood flow due to the release of endogenous histamine or prostaglandins. Tracer [¹²⁵I]CGRP (100 nCi, 0.05 pmol) was injected in a 0.1 ml volume into rat skin at two sites on each rat, alone (●) and with C.48/80 (1 μ g per site, ○). Using an external collimated gamma probe and rate meter, counts at each of the injected sites were measured initially and then every 5 min over the next 20 min. Counts from [¹²⁵I]CGRP were lost more rapidly from sites treated with C.48/80 ($P < 0.05$ at 5 min, $P < 0.01$ at 10, 15 and 20 min, Student's t -test). In identical experiments, where [¹²⁵I]CGRP was replaced by a clearance isotope, ¹³³Xe, to monitor blood flow³, C.48/80 did not significantly influence flow ($3.8 \pm 2.0\%$ increase, mean $\pm$ s.e.m., $n = 4$).



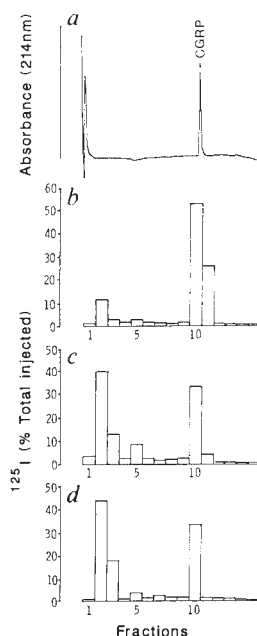
which is dependent on mast cell activation, but independent of histamine release.

To analyse the mechanisms involved, we performed studies in rat skin. Figure 2*a* shows the effect of substance P on the elevated blood flow induced by intradermally injected CGRP. Rats were pretreated systemically with an H₁ anti-histamine and an anti-serotonin (see Fig. 2) to counter the effects of mast-cell-derived amines. No significant change in blood flow was detectable with substance P alone. However, substance P induced a dose-related reduction in the vasodilatation stimulated by CGRP. The attenuating effect of substance P on vasodilatation

was accompanied by local oedema (see Fig. 2*a*) because substance P directly increases vascular permeability, as well as the release of mast cell amines. These oedema responses induced by substance P are, in fact, potentiated by the initial vasodilator activity of CGRP (refs 7 and 14; Fig. 2). To ensure that local oedema formation was not influencing the loss of the vasodilator activity of CGRP, further experiments were carried out using compound 48/80 (C.48/80). C.48/80 activates mast cells, but does not directly increase permeability. C.48/80 did not therefore induce oedema in the rats pretreated with amine antagonists (Fig. 2). Like substance P, C.48/80 also attenuated vasodilata-

Fig. 4 Breakdown of [125 I]CGRP during incubation with mast cells stimulated with substance P (0.2–2 nmol per 0.1 ml) and C.48/80 (0.2 μ g per 0.1 ml) *in vitro*. **a**, Elution of CGRP (0.2 nmol, injected in 20 μ l). Elution of radioactivity from a sample of [125 I]CGRP incubated with **b** unstimulated mast cells, **c** mast cells stimulated with SP (2 nmol per 0.1 ml) and **d** mast cells stimulated with C.48/80. Percentage radioactivity co-eluting with CGRP, in fractions 10 and 11, was as follows: unstimulated mast cells $73.8 \pm 5.0\%$, mean $\pm$ s.e.m., $n = 6$ which was significantly more (Student's *t*-test) than that eluting from samples incubated with mast cells stimulated with SP (0.2 nmol per 0.1 ml), $54 \pm 3.3\%$, $n = 4$ ($P < 0.05$); mast cells stimulated with SP (2 nmol per 0.1 ml), $55.3 \pm 5.9\%$, $n = 6$, ($P < 0.001$); mast cells stimulated with C.48/80 (0.2 μ g per 0.1 ml), $58.9 \pm 6.1\%$, $n = 6$ ($P < 0.01$). Most of the remaining radioactivity eluted in fractions 2 and 3, where tryptic and chymotryptic fragments of CGRP elute. Samples from similar experiments were injected intradermally in the rabbit and blood flow measured at 30 min as in Fig. 2. After incubation with substance-P-stimulated mast cells (0.2 nmol per 0.1 ml), $31.3 \pm 6.1\%$ of the vasodilator activity of CGRP remained when compared with the activity of CGRP measured after incubation with unstimulated mast cells ($P < 0.05$, $n = 5$, Student's *t*-test). An antibody (Ab) to substance P was added to samples after incubation in order to neutralize the substance P. The antibody had no effect on the vasodilator activity of CGRP and no local oedema formation was observed at injected sites in the presence of the antibody to substance P. Oedema (μ l plasma per skin site) was as follows: PBS + Ab, 3.6 ± 1.1 μ l; CGRP + Ab, 5.3 ± 1.2 μ l; SP + CGRP + Ab, 5.3 ± 0.6 μ l. C.48/80 did not, at the concentrations used (0.2 μ g per 0.1 ml), affect the rabbit microvasculature and $37.6 \pm 10.5\%$ of the vasodilator activity remained after incubation of CGRP with C.48/80-stimulated mast cells when compared with the activity of CGRP incubated with unstimulated mast cells ($P < 0.01$, $n = 8$ Student's *t*-test).

Methods. Mixed peritoneal cells containing 1–5% mast cells (assessed by staining with neutral red) were obtained and prepared from male Sprague-Dawley rats¹¹. The cells (total 10^6 per 0.1 ml) were incubated with the agents under test for 5 min at 37 °C. Parallel incubations were carried out with CGRP ([125 I]CGRP, tracer (10 nCi) + CGRP, final concentration 100 pmol per ml) incubated with cells, cells plus activators or buffer alone. All incubations were terminated by cooling to 0 °C and centrifugation (200g for 5 min). Supernatants were diluted with an equal amount of trifluoroacetic acid (0.08%), centrifuged again (10,200g for 3 min) and then frozen. Samples (20 μ l) with 0.2 nmol cold CGRP to act as carrier, were subjected to C18 reversed-phase HPLC on a Vydac 218 TP54 silica analytical column using a linear gradient of 24–60% acetonitrile over 20 min at a flow rate of 1.5 ml min⁻¹. One-minute fractions were collected and radioactivity measured.



tion induced by CGRP (Fig. 2b). These results indicate that the attenuation of CGRP-induced vasodilatation can be dissociated from local oedema formation and also implicate mast cells in the attenuation of responses to CGRP.

Mast cells contain proteases that can be released or expressed on the cell surface in response to stimulation^{15,16}. Recently it has been shown that trypsin and chymotrypsin, purified from dog mastocytoma, break down neuropeptides *in vitro*¹⁷. Figure 2b shows that the attenuation of CGRP-induced vasodilatation by intradermally injected C.48/80 was prevented by the addition of the protease inhibitor soya bean trypsin inhibitor (SBTI). Figure 2c shows that pre-incubation of trypsin or chymotrypsin with CGRP *in vitro*, before intradermal injection, also suppressed vasodilator activity, and this was also prevented by SBTI.

One route of removal of peptides in a tissue is clearance through the bloodstream. Diffusion rate into blood micro-vessels is inversely related to peptide size. We therefore thought that degraded CGRP would be removed faster than the intact peptide. We compared the clearance of [125 I]CGRP from rat skin, with and without the addition of C.48/80. The rats were treated with indomethacin and cimetidine (in addition to mepyramine and methysergide, Fig. 2), to ensure that there was no vasodilator contribution by any endogenously generated prostaglandins or histamine; no detectable increased blood flow in response to C.48/80 was observed (Fig. 3). There was, however, a marked increase in the clearance of radioactivity in the presence of C.48/80 as shown in Fig. 3.

To support the hypothesis that mast cells can be involved in CGRP metabolism *in vivo*, we carried out experiments with isolated mast cells *in vitro*. CGRP, with [125 I]CGRP tracer, was incubated at 37 °C for 5 minutes with rat peritoneal mast cells in the presence and absence of either substance P or C.48/80. The samples were fractionated by high-performance liquid chromatography (HPLC). Unstimulated cells had little effect on [125 I]CGRP, whereas cells in the presence of either substance P or C.48/80 caused a loss of radioactivity co-eluting with intact CGRP (Fig. 4). This loss corresponded to an increase in radioactivity eluting earlier, where fragments of CGRP elute. Vasodilator activity was also lost after incubation of CGRP with activated mast cells (Fig. 4).

These results indicate that a new mechanism may operate to inactivate certain neuropeptides *in vivo*. If both CGRP and substance P are released simultaneously, CGRP can exert its potent relaxant effect on vascular smooth muscle cells. The increased local blood flow can then be controlled by CGRP proteolysis by enzymes released from, or expressed on, mast cells in response to substance P. This inactivating mechanism may be outweighed by an excess of CGRP generated locally, by local depletion of substance P, or by mast cell protease depletion. These events could contribute to the protracted erythema seen in response to local infection or injury. As well as being a new mechanism for the inactivation of a neuropeptide this is the first time that the attenuation of a biological effect has been shown to occur *in vivo* as a direct consequence of mast cell activation.

This investigation was supported by the Halley Stewart Trust and the Medical Research Council. We are grateful for additional support from Merck Sharp and Dohme Research Laboratories, Rahway, and Boots PLC, Nottingham. We thank Dr P. J. Jose and Mr S. B. Coade for helpful advice. We are indebted to the League of Friends, Northwick Park Hospital who purchased the laser-Doppler blood flow meter used in these studies.

Received 24 February; accepted 12 July 1988.

- Rosenfeld, M. G. *et al.* *Nature* **304**, 129–135 (1983).
- Brain, S. D., Williams, T. W., Tippins, J. R., Morris, H. R. & MacIntyre, I. *Nature* **313**, 54–56 (1985).
- Mulderry, P. K. *et al.* *Neuroscience* **14**, 947–954 (1985).
- Brain, S. D., Tippins, J. R., Morris, H. R., MacIntyre, I. & Williams, T. J. *J. invest. Derm.* **87**, 533–536 (1986).
- Gibbins, I. L. *et al.* *Neurosci. Lett.* **57**, 125–130 (1985).
- Lundberg, J. M., Franco-Cereceda, A., Hua, X.-Y., Hokfelt, T. & Fischer, J. A. *Eur. J. Pharmacol.* **108**, 315–319 (1985).
- Brain, S. D. & Williams, T. J. *Br. J. Pharmac.* **86**, 855–860 (1985).
- Dale, H. H. *Bull. Johns Hopkins Hosp.* **53**, 297–347 (1933).
- Burnstock, G. *Neuroscience* **1**, 239–248 (1976).
- Hagermark, O., Hokfelt, T. & Pernow, B. *J. invest. Derm.* **71**, 233–235 (1978).
- Foreman, J. C., Jordan, C. C., Oehme, P. & Renner, H. J. *Physiol.* **335**, 449–465 (1983).
- Devillier, P. *et al.* *Pharmacology* **32**, 340–347 (1986).
- Fuller, R. W., Conradson, T.-B., Dixon, C. M. S., Crossman, D. C. & Barnes, P. J. *Br. J. Pharmac.* **92**, 781–788 (1987).
- Gamse, R. & Saria, A. *Eur. J. Pharmacol.* **114**, 61–66 (1985).
- Lagonoff, D. in *Biochemistry of the Acute Allergic Reactions* (eds Becker, E., Simon, A. & Austen, K. F.) 89–101 (Liss, New York, 1981).
- Gervasoni, J. E., Conrad, D. H., Hugli, T. E., Schwartz, L. B. & Ruddy, S. J. *Immun.* **136**, 285–292 (1986).
- Caughy, G. H., Leidig, F., Viro, N. F. & Nadel, J. A. *J. Pharmac. exp. Ther.* **244**, 133–137 (1988).
- Holloway, G. A. & Watkins, D. M. *J. invest. Derm.* **69**, 306–309 (1977).

Structural and functional basis for GABA_A receptor heterogeneity

Edwin S. Levitan*†, Peter R. Schofield†‡§, David R. Burt*†, Lucy M. Rhee§, William Wisden*, Martin Köhler‡, Norihisa Fujita*†, Henry F. Rodriguez§, Anne Stephenson*, Mark G. Darlison*, Eric A. Barnard* & Peter H. Seeburg‡§||

* MRC Molecular Neurobiology Unit, MRC Centre, Hills Road, Cambridge CB2 2QH, UK

‡ Laboratory of Molecular Neuroendocrinology, ZMBH, University of Heidelberg, Im Neuenheimer Feld 282, 6900 Heidelberg, FRG

§ Genentech, Inc., Department of Developmental Biology, 460 Point San Bruno Boulevard, South San Francisco, California 94080, USA

When γ -aminobutyric acid (GABA), the major inhibitory neurotransmitter in vertebrate brain, binds to its receptor it activates a chloride channel. Neurotransmitter action at the GABA_A receptor is potentiated by both benzodiazepines and barbiturates which are therapeutically useful drugs (reviewed in ref. 1). There is strong evidence that this receptor is heterogeneous¹⁻⁷. We have previously isolated complementary DNAs encoding an α - and a β -subunit and shown that both are needed for expression of a functional GABA_A receptor⁸. We have now isolated cDNAs encoding two additional GABA_A receptor α -subunits, confirming the heterogeneous nature of the receptor/chloride channel complex and demonstrating a molecular basis for it. These α -subunits are differentially expressed within the CNS and produce, when expressed with the β -subunit in *Xenopus* oocytes, receptor subtypes which can be distinguished by their apparent sensitivity to GABA. Highly homologous receptor subtypes which differ functionally seem to be a common feature of brain receptors.

Of 12 bovine brain α -subunit cDNA clones isolated⁸, one contained an open reading frame encoding a different 451 amino acid polypeptide (Fig. 1). The first 28 amino acids of this polypeptide λ have typical features of a signal peptide⁹ and the predicted relative molecular mass (M_r) of the mature polypeptide is 48,000 (48 K). The corresponding amino acid sequence shows 79% identity with the previously characterized GABA_A receptor α -subunit⁸, but only 34% amino acid identity with the mature GABA_A receptor β -subunit⁸. We therefore called the subunit that this cDNA encodes an $\alpha 2$ subunit, with the previously characterized clone encoding an $\alpha 1$ subunit. The predicted $\alpha 2$ subunit polypeptide sequence also contained a peptide sequence that was chemically determined from the affinity-purified adult GABA_A receptor⁸, indicating that the new subunit is present in this receptor (Fig. 1). Re-screening of the bovine brain cDNA library revealed the existence of additional cDNAs encoding α -subunits. Ten cDNAs were analysed and the longest three sequenced. These all contained putative unspliced introns at two different locations but had the potential (after intron removal) to encode a polypeptide of 492 amino acids (Fig. 1). A 28-residue signal peptide⁹ and a mature polypeptide of M_r 52 K is predicted. This subunit, designated $\alpha 3$, displays 72% sequence identity to both the $\alpha 1$ and $\alpha 2$ subunits. The receptor heterogeneity observed arises from different genes encoding the respective subunits and not by alternative splicing.

The presence of four hydrophobic transmembrane domains, M1-M4, deduced from both the $\alpha 1$ and β chains are present

at equivalent positions in the $\alpha 2$ and $\alpha 3$ polypeptides⁶. The sequences of M1-M3 are extremely highly conserved in all three α -subunits with only a single conservative amino acid change ($I > M_{314}$ in the $\alpha 3$ sequence) over a stretch of 98 amino acids (Fig. 1). The M4 region is also highly conserved, as is the majority of the presumed extracellular large N-terminal domain, which shows 75% sequence identity in all three chains (Fig. 1). The region between M3 and M4, deduced to be intracellular, and the extreme N-terminal sequences (including the signal peptides) are very poorly conserved. Each α -subunit contains potential N-linked glycosylation sites: two in $\alpha 1$, three in $\alpha 2$ and four in $\alpha 3$. The hypothesis that members of the ligand-gated ion channel superfamily have a common evolutionary origin^{8,10} is strengthened by the homologies seen here and by the homologous positions of some of the putative unspliced intron sequences in the $\alpha 3$ subunit and the α , γ and δ subunits of the nicotinic acetylcholine receptor genes¹¹⁻¹³.

We investigated whether the three α -subunits confer different properties on the GABA_A receptor by expressing each α RNA in combination with the β -subunit RNA in *Xenopus* oocytes. Electrophysiological recording showed that either the $\alpha 2$ or the $\alpha 3$ polypeptide could be substituted for the $\alpha 1$ polypeptide to yield GABA-evoked currents. These were always inhibited similarly by bicuculline and picrotoxin (Fig. 2a). The barbiturate pentobarbital directly activated each receptor and markedly enhanced the responses to GABA (Fig. 2b). Plots of current against voltage demonstrated that receptors containing any one of the three α -subunit polypeptides had reversal potentials (E_r) close to the chloride equilibrium potential¹⁴ ($E_r \approx -20$ mV, Fig. 2c), indicating that each receptor subtype forms GABA-gated chloride channels. Finally, the currents due to the ($\alpha 2 + \beta$) and the ($\alpha 3 + \beta$) receptors do not increase linearly with voltage, but show marked outward rectification (Fig. 2c). Identical voltage-dependence is seen with the ($\alpha 1 + \beta$) receptor (data not shown). Thus, substitution of either the $\alpha 2$ or the $\alpha 3$ subunit for the $\alpha 1$ subunit does not appear to change the qualitative properties of the expressed GABA_A receptor.

Exchanging the α -subunits, however, alters the apparent sensitivity of the receptor to neurotransmitter (Fig. 3). The half-maximal dose for GABA varies significantly between receptor subtypes: 1.3 ± 0.2 μ M ($n = 6$ oocytes) for ($\alpha 2 + \beta$), 12 ± 1 μ M ($n = 6$) for ($\alpha 1 + \beta$) and 42 ± 8 μ M ($n = 5$) for ($\alpha 3 + \beta$). The differences in GABA sensitivity could be due to altered receptor affinities, changes in coupling of the receptor to channel gating, or desensitization kinetics. The latter seems less likely since the differences in GABA sensitivity were consistent, despite large variations in desensitization. Moreover, for ($\alpha 1 + \beta$) and ($\alpha 3 + \beta$), Hill plots were linear (data not shown). Our results demonstrate that exchanging the α -subunits can produce an apparent 30-fold change in sensitivity of GABA_A receptors expressed in oocytes.

The expressed receptors display differences in their apparent sensitivity to GABA which are indicative of receptor subtypes, and share many properties with native GABA_A receptors. These include competitive inhibition by bicuculline, voltage-dependent enhancement and direct activation by pentobarbital, anion selectivity, voltage-dependent gating (data not shown) and characteristic multiple single-channel conductance states (L. Blair and V. Dionne, unpublished). The expressed receptors are, however, not identical with native GABA_A receptors. Hill coefficients derived from the data in Fig. 3 are less than or equal to 1, rather than 1.4 to 2 as expected^{15,16}. Benzodiazepine potentiation was not observed, whereas a modest response in follicle-enclosed oocytes to a single benzodiazepine had been seen previously⁸. Neither co-expression of all four known subunits nor the use of a range of benzodiazepines gave the receptor potentiation observed when brain poly(A)⁺ RNA is expressed in *Xenopus* oocytes¹⁶⁻¹⁸. Additionally, transiently transfected mammalian cells expressing ($\alpha 1 + \beta$) subunits do not bind benzodiazepines (Pritchett *et al.*, submitted). Specific post-transla-

† Present addresses: Department of Pharmacology, Yale University School of Medicine, New Haven, Connecticut 06510, USA (E.S.L.); Pacific Biotechnology Ltd, 72 McLachlan Avenue, Rushcutters Bay 2011, Australia (P.R.S.); Department of Pharmacology, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA (D.R.B.); Biorganic Research Department, CIBA-GEIGY (Japan) Ltd, Takarazuka 665, Japan (N.F.).

|| To whom correspondence should be addressed at Heidelberg.

The three α -subunits of the GABA_A receptor differ in both

Fig. 2 Expression of cloned GABA_A receptors in oocytes injected with ($\alpha 2 + \beta$) (left) or ($\alpha 3 + \beta$) (right) RNAs. All data are from defolliculated oocytes, except the ($\alpha 3 + \beta$) traces in *a* and *b*. *a*, Effects of bicuculline methobromide and picrotoxin on GABA-evoked currents. GABA applications (0.2 μ M for $\alpha 2 + \beta$; 2 μ M for $\alpha 3 + \beta$) are indicated by bars. The different concentrations were used to compensate for the different sensitivities (Fig. 3). Bicuculline (BIC), 4 μ M, or picrotoxin (PTX), 2.5 μ M, was applied for 1.5 minutes; GABA was then applied in the continued presence of each drug. The traces are from the same oocyte with the bicuculline washout not shown. *b*, Effects of the barbiturate pentobarbital. 20 μ M pentobarbital (PB) directly activates the expressed receptors (no effect is seen in uninjected oocytes) and enhances the responses to GABA (0.1 μ M for $\alpha 2 + \beta$, 1 μ M for $\alpha 3 + \beta$). *c*, Current/voltage relationships for the GABA-evoked current. Measurements were made during a voltage ramp (300 mV s⁻¹). Currents before application of GABA (1 μ M for ($\alpha 2 + \beta$); 30 μ M for ($\alpha 3 + \beta$)) were subtracted. The reversal potential is the voltage where zero current flows through the membrane. Reversal potentials (E_r) were: ($\alpha 2 + \beta$) $E_r = -24 \pm 2$ mV ($\pm$ standard error of the mean, $n = 4$ oocytes); ($\alpha 3 + \beta$) $E_r = -19 \pm 2$ mV ($n = 6$). Similar results as *a*, *b* and *c* were also obtained with oocytes injected with ($\alpha 1 + \beta$), $E_r = -20 \pm 1$ mV ($n = 6$) (data not shown, see also ref. 8).

Methods. The $\alpha 1$ and β -subunit RNAs were prepared using SP6 RNA polymerase as previously described⁸. The $\alpha 2$ -subunit cDNA was subcloned as a 1.8 kb *PstI*-*SacI* fragment (restriction sites from the polylinker of the sequencing vector M13tg130) into *PstI*-*SacI*-cleaved pSP64 (PbGR $\alpha 2$ sense). The fragment contains the first 1,803 base pairs (bp) of the cDNA sequence. These capped RNA transcripts co-injected with β -subunit RNA produced poor GABA responses, presumably due to the very short 5' untranslated sequence. Accordingly, synthetic sequences were inserted providing 5' untranslated sequences based on the $\alpha 1$ subunit sequences and a Kozak consensus sequence²⁹. The *PstI* (polylinker)-*Bam*HI (nucleotide 68) fragment of pbGR $\alpha 2$ sense was replaced by the synthetic DNA sequence:

```

5'      GAAGCCCACTGAAGACGAACTGAACAGCTCCAATATGCAGTTGCTGCTTTTGTCTTTCTTGGGCTGG      3'
3'      ACGTCTTCGGGTGGTACTTCTGCTTTGACTTGTGCGAGTTATACGTCAACGACGAAAAACAAAGAACCGCACCTAG      5'

```

(The initiation codon is overlined). This plasmid (pbGR $\alpha 2$ sense + K) was linearized with *PstI* and the synthetic DNA sequence:

```

5'      GGATTTTCTCTGCGAGACTTTTCCCGGGTCTGGAGCGATCCTGTGCCAGAGGGGGCCGAGCTGCA      3'
3'      ACGTCTCTAAAGAGAGCGTCTGAAAAGGGCCGACCTCGTAGGACACGGGTCTCCCCGGGCTCG      5'

```

inserted in the orientation shown. After linearization of the resultant plasmid (pbGR $\alpha 2$ sense + K + 5'ut) with *SacI*, capped SP6 polymerase transcripts were obtained. These gave the GABA responses shown. For the $\alpha 3$ subunit, the cDNA λ bGR9.3 was subcloned as a 3.6 kb *XhoI* (restriction sites in cDNA adaptors) fragment into *SalI* cleaved pBS+ (Stratagene). A ~2.0-kb *NcoI* fragment beginning at nucleotide 1,069 was replaced with the equivalent fragment from λ bGR1. This allowed the removal of the unspliced intron of the former clone. After linearization of pbGR $\alpha 3$ with *HindIII*, capped RNA transcripts were obtained using T7 RNA polymerase. Other details were as in ref. 8. Each RNA (10 ng) was injected into *Xenopus* oocytes (stages 5 and 6). Oocytes were defolliculated manually after 15–90 min in 2 mg ml⁻¹ collagenase in Ca²⁺-free modified Barth's medium³⁰, either before or 1 day after the RNA injection. After 2–10 days of incubation in modified Barth's medium supplemented with 2.5 mM pyruvate³⁰, oocytes were placed in a 120 μ l oblong well, impaled with 2 electrodes filled with 3M KCl and voltage clamped at -60 mV. The preparation was superfused with saline at 6 ml min⁻¹ and GABA was bath-applied with a dead-time of 4 s. No qualitative differences in the GABA responses shown were observed with and without defolliculation. All experiments were performed on 3–6 oocytes, with consistent results.

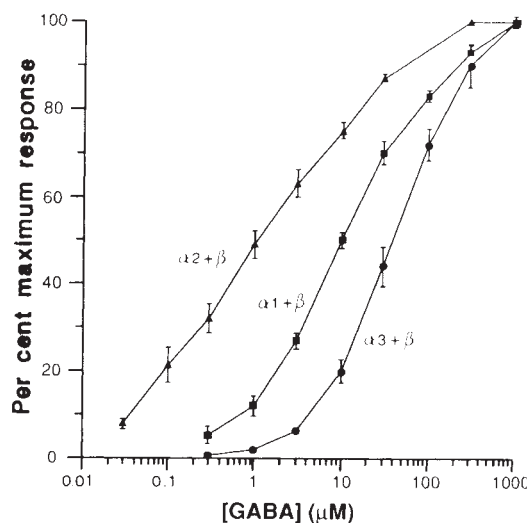


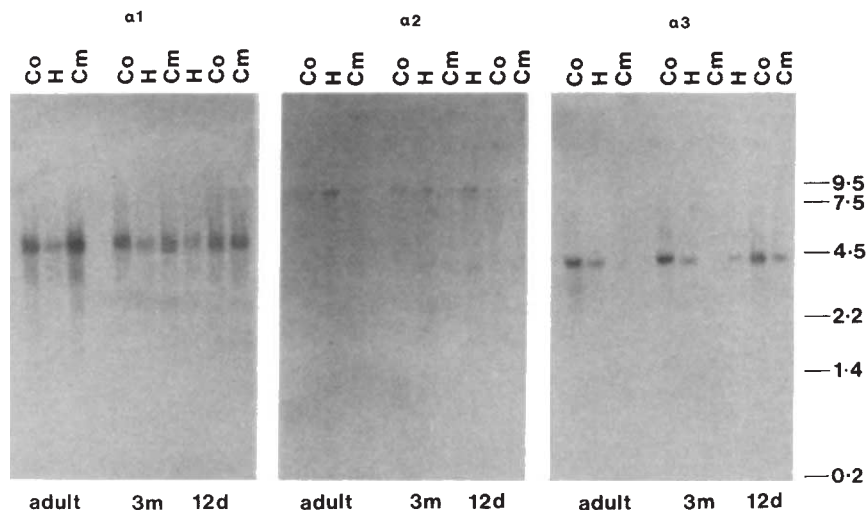
Fig. 3 Exchanging α -subunits alters the sensitivity of the receptor to GABA. Dose-response curves from oocytes injected with ($\alpha 1 + \beta$), ($\alpha 2 + \beta$) or ($\alpha 3 + \beta$) were constructed using the peak responses, to minimize the effects of desensitization. The data for ($\alpha 1 + \beta$) and ($\alpha 3 + \beta$) were normalized to the response to 1 mM GABA. ($\alpha 2 + \beta$) data were normalized to the 300 μ M response. Each point is the mean from five or six defolliculated oocytes from at least two donors, voltage-clamped at -60 mV. The half-maximal doses were obtained from Hill plots of these data. The half-maximal dose for ($\alpha 1 + \beta$) was significantly different from those for ($\alpha 2 + \beta$) ($P < 0.001$) and ($\alpha 3 + \beta$) ($P < 0.01$). Maximal responses in Cl⁻ conductance were: 7.45 ± 2.40 μ S ($n = 6$) for ($\alpha 1 + \beta$); 13.8 ± 4.7 μ S ($n = 6$) for ($\alpha 2 + \beta$); 67.2 ± 27.9 μ S ($n = 5$) for ($\alpha 3 + \beta$).

size and sequence. Significantly, they confer different apparent sensitivities to GABA in the expressed receptors. Pharmacological alteration of GABA_A receptor sensitivity has been shown to change both the size and duration of GABAergic synaptic potentials *in vitro*²⁴. It is therefore likely that neuronal responsiveness to GABAergic input is influenced by differential expression of

α -subunit genes. Northern blot analysis reveals that this expression varies between different brain regions and possibly during development. Thus, differential expression of multiple receptor subtype genes may be used to generate a greater diversity of GABA responses than would be achieved with a single subtype. Such differential expression of homologous but func-

Fig. 4 Regional brain distribution of α -subunit mRNAs. Northern blots of poly(A)⁺ RNA from adult, 3-month-old (3m) and 12-day-old (12d) bovine cortex (Co), hippocampus (H) and cerebellum (Cm) were probed with α 1-, α 2- and α 3-specific oligonucleotide probes.

Methods. Total RNA was isolated by a modification of the guanidinium/hot phenol method³¹ and poly(A)⁺ RNA was extracted by oligo(dT)-cellulose chromatography. Poly(A)⁺ RNA (25 μ g of each preparation) was electrophoresed in a 1% (w/v) formaldehyde agarose gel, blotted onto a nylon membrane (Hybond-N, Amersham) and UV cross-linked. Probes used were subunit-specific oligonucleotides: α 1-specific 45-mer (5'-GGGTCACCCCTGGCCA-GATTAGGTGTGTAGCTGGTTGCTGTG-GA3'), α 2-specific 45-mer (5'-ACTGGGTCTT-TTGAAGATTCGGGGCATAATTGGCAAC-AGCCACT-3') and α 3-specific 45-mer (5'-GGT-GGTTCCACGATGTTGAAAGTGGTGCTG-GTTTCTTGGTCCG-3'). Probes were 3' end-labelled with [³²P]dATP using terminal deoxynucleotidyl transferase to a specific activity of 5×10^8 dpm per μ g. Blots were hybridized in 20% (v/v) formamide at 42 °C and washed in 0.2 $\times$ SSC, 0.1% (w/v) SDS at 60 °C, before exposure to X-ray film with an intensifying screen. An RNA ladder (BRL) was used for size markers. Autoradiographs were exposed for either 4 h (α 1 and α 3) or for 24 h (α 2).



tionally distinct receptor subtypes, known already for the G-protein coupled receptors²⁵⁻²⁸ may be an important mechanism that contributes to synaptic plasticity.

We thank Michael Squire for help with DNA sequencing, Jeanette O'Brien and Brian Morris for some SP6 and mRNA samples, Brenda Shivers for tissue samples, Drs M. Goedert, I. Martin, L. Blair, V. Dionne and B. Shivers for advice and valuable discussions and Jutta Rami for preparation of the manuscript. E.S.L. is supported by the Huntington's Disease Society of America and NSF grant INT-8602995, D.R.B. held a Fogarty Senior International Fellowship, and F.A.S. holds a Royal Society University Research Fellowship. In part, this work was supported by a grant from the Deutsche Forschungsgemeinschaft to P.H.S. Authors E.S.L. and P.R.S. are equal first authors.

Received 21 March; accepted 19 July 1988.

A highly conserved amino-acid sequence in thrombospondin, properdin and in proteins from sporozoites and blood stages of a human malaria parasite

Kathryn J. H. Robson, Jennifer R. S. Hall, M. W. Jennings, T. J. R. Harris*, K. Marsh, C. I. Newbold, Valerie E. Tate† & D. J. Weatherall

MRC Molecular Haematology and Tropical Medicine Units, Nuffield Department of Clinical Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK

* Molecular Biology, Celltech Limited, 216 Bath Road, Slough, UK

As a consequence of gene cloning and DNA sequencing several gene families are emerging in the field of cell-cell recognition. These include immunoglobulins, integrins, certain extracellular glycoproteins¹ and a family of functionally unrelated proteins which include factor B². We report here the cloning and sequencing of a gene from *Plasmodium falciparum*, coding for a protein we call thrombospondin related anonymous protein (TRAP), which shares certain sequence motifs common to other well-characterized proteins. The most significant homology is based around the sequence Trp-Ser-Pro-Cys-Ser-Val-Thr-Cys-Gly (WSPCSVTCG), present in three copies in region I of thrombospondin (TSP)³, six copies in properdin⁴ (P) and one copy in all the circumsporozoite (CS) proteins⁵⁻¹⁰ sequenced so far. TRAP also shares with certain extracellular glycoproteins, including TSP, the cell-recognition signal Arg-Gly-Asp (RGD)¹¹, which has been shown to be crucial in the interaction of several extracellular glycoproteins with members of the integrin superfamily. Unlike the CS protein, TRAP is expressed during the erythrocytic stage of the parasite life cycle.

The human malaria parasite, *P. falciparum* follows a complex life-cycle, during which different antigens are expressed at particular developmental stages. The major antigen on the sporozoite surface is the CS protein, which probably determines the specificity of the interaction between the parasite and liver cells. CS proteins contain two conserved amino-acid sequences,

- Olsen, R. W. & Venter, J. C. (eds) *Benzodiazepine/GABA Receptors and Chloride Channels: Structural and Functional Properties*. (Liss, New York, 1986).
- Braestrup, C. & Nielsen, M. J. *J. Neurochem.* **37**, 333-341 (1981).
- Squires, R. F. *et al. Pharmac. biochem. Behav.* **10**, 825-830 (1979).
- Lippa, A. S. *et al. Brain Res. Bull.* **14**, 189-195 (1985).
- Cooper, S. J., Karkham, T. C. & Estall, L. B. *Trends pharmac. Sci.* **8**, 180-184 (1987).
- Akaike, N., Inoue, M. & Krishtal, O. A. *J. Physiol. Lond.* **379**, 171-185 (1986).
- Cash, D. J. & Subbarao, K. *Biochemistry* **26**, 7562-7570 (1987).
- Schofield, P. R. *et al. Nature* **328**, 221-227 (1987).
- von Heijne, G. *Nucleic Acids Res.* **14**, 4683-4690 (1986).
- Grenningloh, G. *et al. Nature* **328**, 215-220 (1987).
- Noda, M. *et al. Nature* **305**, 818-823 (1983).
- Nef, P., Mauron, A., Stalder, R., Alliod, C. & Ballivet, M. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7975-7979 (1984).
- Shibahara, S. *et al. Eur. J. Biochem.* **146**, 15-22 (1985).
- Kusano, K., Miledi, R. & Stinnakre, J. *J. Physiol.* **328**, 143-170 (1982).
- Smart, T. G., Houamed, K. M., Van Renterghem, C. & Constanti, A. *Biochem. Soc. Trans.* **15**, 117-122 (1987).
- Sigel, E. & Bauer, R. *J. Neurosci.* **8**, 289-298 (1988).
- Smart, T. G., Constanti, A., Bilbe, G., Brown, D. A. & Barnard, E. A. *Neurosci. Lett.* **40**, 55-59 (1983).
- Houamed, K. M. *et al. Nature* **310**, 318-321 (1984).
- Kuriyama, K. & Taguchi, J. *J. Neurochem.* **48**, 1897-1903 (1987).
- Sweetnam, P. & Tallman, J. F. *Molec. Pharmac.* **29**, 299-306 (1986).
- Study, R. E. & Barker, J. L. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7180-7184 (1981).
- Unnerstall, J. R., Kuhar, M. J., Niehoff, D. L. & Palacios, J. M. *J. Pharm. exp. Ther.* **218**, 797-804 (1981).
- de Blas, A. L., Vitorica, J. & Friedrich, P. *J. Neurosci.* **8**, 602-614 (1988).
- Segal, M. & Barker, J. L. *J. Neurophysiol.* **52**, 469-487 (1984).
- Nathans, J., Thomas, D. & Hogness, D. S. *Science* **232**, 193-202 (1986).
- Bonner, T. I., Buckley, N. J., Young, A. C. & Brann, M. R. *Science* **237**, 527-532 (1987).
- Peralta, E. G. *et al. EMBO J.* **6**, 3923-3929 (1987).
- Frielle, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 7920-7924 (1987).
- Kozak, M. *Cell* **44**, 283-292 (1986).
- Barnard, E. A. & Bilbe, G. in *Neurochemistry: A Practical Approach* (eds Turner, A. J. & Bachelard, H.) 243-270 (IRL, Oxford, 1987).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning—A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).

† Present address: Cancer Research Campaign Developmental Tumours Research Group, Department of Zoology, South Parks Road, Oxford OX1 3PF, UK.

Methods. The 27-base oligonucleotide sequence was synthesized using an Applied Biosystems model 308A synthesizer by monomer addition of phosphoramidates to a solid support. The complementary strand was also synthesized so that both RNA and DNA sequences could be detected. The de-protected probe was purified by preparative polyacrylamide gel electrophoresis. The end-labelled oligonucleotide was used to screen 4×10^5 plaques of an *EcoRI* digest of T9/96 DNA cloned into the *EcoRI* site of λ gt11. These contained 2.35-kb *EcoRI* fragments. Hybridization was carried out in 0.5 M NaCl, 0.01 M EDTA, 5 $\times$ Denhardt's, 0.5% NP40, 100 μ g ml⁻¹ (1 $\times$ SSC = 0.15 M NaCl, 0.015 M sodium citrate) 0.1% SDS, followed by autoradiography for 16 h with pre-flashed X-ray film.

CS protein	Amino acid sequence	P	C	S	V	T	C	G	K	G
	DNA sequence	CCA	TGT	AGT	GTA	ACT	TGT	GGA	AAT	GGT
TRAP	Amino acid sequence	P <th>C</th> <th>S</th> <th>V</th> <th>T</th> <th>C</th> <th>G</th> <th>K</th> <th>G</th>	C	S	V	T	C	G	K	G
	DNA sequence	CCA <td>TGT</td> <td>AGT</td> <td>GTA</td> <td>ACT</td> <td>TGT</td> <td>GGA</td> <td>AAA</td> <td>GGT</td>	TGT	AGT	GTA	ACT	TGT	GGA	AAA	GGT

[illegible]

Methods. The 2.35-kb *EcoRI* insert from the λ gt11 phage was subcloned into pUC8 and M13mp10. Dideoxy sequencing of this fragment revealed an incomplete open reading frame. A second library was constructed by cloning a partial *EcoRI* digest of T9/96 DNA into λ gt10, and was screened with the original 2.35-kb fragment. Clones with overlapping sequences were obtained. The complete DNA sequence of the TRAP gene and its flanking sequences was established using the chain termination method of Sanger *et al.*²¹. The information was obtained from 'shotgun' clones in M13mp10 which enabled both DNA strands to be sequenced. Gaps were filled in using custom-synthesized oligonucleotide primers. This approach proved the most efficient way of deriving the DNA sequence where there was an absence of restriction enzyme sites; a particular problem in regions of A-T-rich DNA. The DNA sequence handling programmes of Staden were used to analyse the data²². The site to which the oligonucleotide probe hybridized is underlined. The asparagine residues which are potential sites for N-linked glycosylation have asterisks. The RGD sequence (residues 307–309) is over-lined. The IQQ motif (residues 76–78) is boxed.

To search for CS protein-related sequences in the genome of *P. falciparum* the oligonucleotide probe shown in Fig. 1a, corresponding to region II, was used to probe a genomic Southern blot of *P. falciparum* (T9/96) DNA. The predicted 9-kilobase (kb) *Eco*RI and 800-base-pair (bp) *Bst*NI fragments were de-

tested. The same probe was used to screen two genomic *P. falciparum* DNA libraries, one a complete *Eco*RI digest cloned in λ gt11 and the other a partial *Eco*RI digest cloned in λ gt10. The true CS protein sequence was not expected to be found in either of these libraries because the two vectors have an upper limit of 8 kb. Several clones were isolated all of which shared a 2.35-kb *Eco*RI fragment. This fragment and that of a neigh-

Fig. 3 Comparison of amino-acid sequences around and including the conserved nonapeptide motif. The TRAP sequence (residues 244–292) is compared with the equivalent regions of the CS proteins from *P. falciparum* (P.f.)⁵, *P. vivax* (P.v.)⁶, *P. knowlesi* (P.k.)⁷, *P. cynomolgi* London strain (P.c.)⁸, *P. berghei* (P.b.)⁹, *P. yoelii* (P.y.)¹⁰, thrombospondin³ and the properdin framework⁴ using the ALIGN program²³. The one-letter amino-acid code has been used. Residues in common with TRAP are boxed.

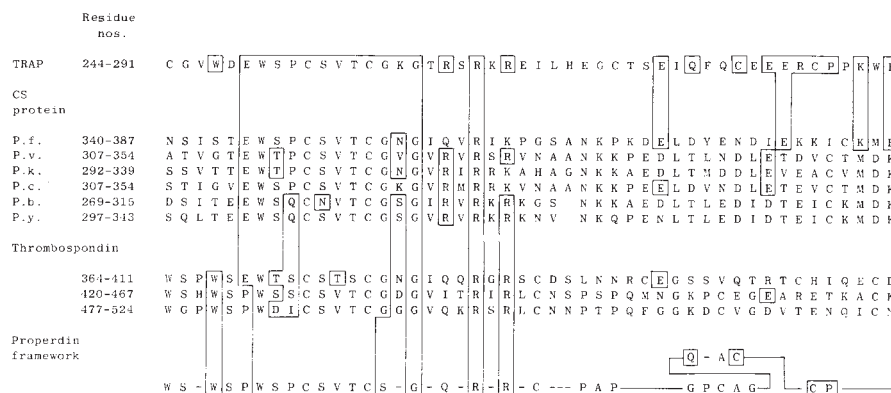


Fig. 4 Southern blot analysis of DNA from the cloned line T9/96 of *P. falciparum* hybridized with TRAP (a) and CS protein (b)²⁴ probes. Lanes 1-11 correspond to restriction digests with the following enzymes, *Asp* 718, *Bam*HI, *Bgl*II, *Eco*RI, *Hinc* II, *Hind*III, *Taq*I, *Bst*NI, *Hha*II and *Msp*I respectively. c, Northern blot analysis of total RNA from: lane 1, EBV transformed lymphocytes (25 µg); lane 2, *P. falciparum* isolate ITO (25 µg); lane 3, *P. falciparum* FCR3 A2 (cloned) (25 µg).

Methods. Southern blotting and hybridization was carried out as described by Woo *et al.*²⁵. RNA extraction²⁶ and Northern blotting was as described by Robson *et al.*²⁷.

Hybond-N membranes were used to permit re-hybridization of the same filter. The filter in panel A was initially hybridized with the two *Eco*RI fragments corresponding to the TRAP sequence; the signal was removed and the filter rehybridized with the probe containing the CS protein gene sequence. The probe panel C was the same as panel A. The filter in panel C was also hybridized to the same probe as panel B; no hybridization signal was seen.

bouring *Eco*RI fragment were sequenced (Fig. 2), identifying an open reading frame of 559 amino acids, predicted to code for a protein of relative molecular mass (M_r) 63,300. The amino-acid sequence includes the conserved nonapeptide (Trp-Ser-Pro-Cys-Ser-Val-Thr-Cys-Gly), explaining why the oligonucleotide probe had detected this new gene (Fig. 1*b*, ref. 5). This sequence and variations of it are found in thrombospondin³, CS protein⁵⁻¹⁰ and properdin⁴ (Fig. 3).

The deduced protein sequence has two hydrophobic domains at either end of the molecule. The first, at the amino terminal end is probably a signal sequence; the second, near the carboxy terminus, resembles a transmembrane sequence. There is a cluster of cysteine residues around and including the conserved amino-acid sequence suggestive of a secondary structure formed either by intermolecular or intramolecular disulphide bonds. Cysteine residues occur in similar positions around the conserved regions in CS protein, thrombospondin and properdin. Evidence for such secondary structure comes from Ballou *et al.*¹² who found that antibodies against a CS-derived peptide containing region II reacted poorly with both native and denatured CS protein, suggesting a highly ordered configuration. Beyond the conserved region the sequence becomes rich in proline but this does not form part of a repeat characteristic of many other malaria proteins. Submerged within this sequence is an RGD motif (amino acids 307-309), which is characteristic of many adhesive glycoproteins involved in cell recognition; TRAP is the first malarial protein with this sequence. Thrombospondin has such an RGD motif as well as an IQQ motif which has been implicated in cross-linking to factor XIII_a (ref. 13):

TRAP also has an IQQ motif (amino acids 76-78). There are four possible sites for *N*-glycosylation. Like most malarial antigens the amino-acid composition is unusual in that it is particularly rich in asparagine and proline.

A Southern blot of different restriction enzyme digests of *P. falciparum* DNA was probed sequentially with the TRAP gene and the CS protein gene to determine whether these sequences are independent (Fig. 4a, b). Of the restriction fragments detected by both probes several seemed to be the same size. This may be fortuitous or it may reflect linkage of the two genes.

The CS protein gene is only expressed on sporozoites. Northern blot analysis using a CS gene probe did not detect any RNA species from erythrocytic stages, whereas the TRAP gene probe detected RNA species of about 20S in the two isolates examined, ITO and FCR3A2, but not in EBV-transformed lymphocytes (Fig. 4c). This indicates that the TRAP gene is expressed during the erythrocytic stage of the life cycle. The size of the RNA transcript is compatible with it coding for a protein of M_r 63,300. We have raised antibodies to TRAP β -galactosidase fusion proteins. Preliminary data, from indirect immunofluorescence (K.J.H.R., J.R.S.H., J. Tansey and K.M.) suggest that TRAP is synthesized by mature asexual blood stages.

The occurrence in both vertebrate and invertebrate stages of *P. falciparum* of a highly conserved motif which is also present in thrombospondin and properdin suggests that TRAP might be of functional significance. A possible role for the CS protein of sporozoites is recognition and entry into hepatocytes. Synthetic peptides from region I bind specifically to hepatocytes *in vitro*¹⁴; such studies have not yet been reported for region II.

Two other parasite-cell interactions are critical in the life cycle of *P. falciparum*. Its virulence is related to its propensity to sequester in deep vascular beds reflecting the interaction of parasite-induced modifications on the red cell surface with receptors on endothelial cells. Both thrombospondin itself and thrombospondin antibodies inhibit the cytoadherence of infected red cells in *in vitro* models¹⁵. This, taken together with the involvement of platelet glycoprotein IV (the thrombospondin receptor) in cytoadherence¹⁶ suggests that there may be a parasite-induced thrombospondin analogue on infected erythrocytes. The cytoadherence antigen Pf EMP I is thought to be 300K (ref. 17); this does not exclude the involvement of TRAP as cytoadherence may involve parasite modification of a host protein and TRAP could fulfil this role.

The other parasite-cell interaction is the recognition and invasion of red cells by merozoites. If TRAP were present on free merozoite surfaces the homology with properdin, which binds to C3b, might play a role in the recognition of C3b or its breakdown products on the red cell surface. The closely related parasite *Babesia rhodiani* has developed a strategy for entering erythrocytes involving C3b (ref. 18). The observation that entry of red cells by *P. falciparum* merozoites does not require serum complement components¹⁹ does not exclude the involvement of complement components already on the red cell surface.

A similar size protein to TRAP occurs in *P. knowlesi*. It is synthesized during the late stages of the erythrocytic cycle, and found on the merozoite surface²⁰. Antibodies to this 66K *P. knowlesi* protein allow attachment of the merozoite but prevent subsequent orientation and invasion. It is possible that TRAP is the *P. falciparum* homologue of this 66K protein.

Whichever of these or other possibilities proves to be the case, the remarkable conservation around and including this non-peptide sequence in four apparently unrelated proteins suggests that the significance of this sequence will not be trivial.

We thank Dr S. L. Thein for initial help with the oligonucleotide technology, Miss J. Brown and Mr M. Sheppard for synthesising oligonucleotides, Mrs Liz Rose for typing the manuscript, Dr Lockyer for the CS gene probe and Drs B. Morley, M. Stewart, J. Collins, K. Reid, C. Gilks and A. F. Williams for helpful discussions. We acknowledge financial support (C.I.N. and K.M.) from the UNDP/WORLD BANK/WHO Special Programme for Research and training in Tropical Diseases and the Wellcome Trust.

Received 29 April; accepted 8 July 1988.

- Hynes, R. O. *Cell* **48**, 549-554 (1987).
- Reid, K. B. M. *et al. Immun. Today* **7**, 230-234 (1986).
- Lawler, J. & Hynes, R. O. *J. Cell Biol.* **103**, 1635-1648 (1986).
- Goundis, D. & Reid, K. B. M. *Nature* **335**, 82-85 (1988).
- Dame, J. B. *et al. Science* **225**, 593-599 (1984).
- Arnot, D. E. *et al. Science* **230**, 815-818 (1985).
- Ozaki, L. S., Svece, P., Nussenzweig, R. S., Nussenzweig, V. & Godson, G. N. *Cell* **34**, 815-822 (1983).
- Galinski, M. R. *et al. Cell* **48**, 311-319 (1987).
- Eichinger, D. J., Arnot, D. E., Tam, J. P., Nussenzweig, V. & Enea, V. *Molec. cell. Biol.* **6**, 3965-3972 (1986).
- Lal, A. A. *et al. J. biol. Chem.* **262**, 2937-294 (1987).
- Pierschbacher, M. D. & Ruoslahti, E. *Nature* **309**, 30-33 (1984).
- Ballou, W. R. *et al. Science* **228**, 996-999 (1985).
- Bale, M. D., Westrick, L. G. & Mosher, D. F. *J. biol. Chem.* **260**, 7502-7508 (1985).
- Aley, S. B., Bates, M. D., Tam, J. P. & Hollingdale, M. R. *J. exp. Med.* **164**, 1915-1922 (1986).
- Roberts, D. D. *et al. Nature* **318**, 64-66 (1985).
- Barnwell, J. W., Ockenhouse, C. F. & Knowles, I. D. M. *J. Immun.* **135**, 3494-3497 (1985).
- Howard, R. J. *et al. Molec. biochem. Parasit.* **27**, 207-224 (1988).
- Jack, R. M. & Ward, P. A. *J. Immun.* **124**, 1574-1578 (1980).
- Ward, P. A., Sterzel, R. B., Lucia, H. L., Campbell, G. H. & Jack, R. M. *J. Immun.* **126**, 1826-1828 (1981).
- Deans, J. A. & Jean, W. C. *Molec. biochem. Parasit.* **26**, 155-166 (1987).
- Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. *J. molec. Biol.* **143**, 161-178 (1980).
- Staden, R. *Nucleic Acids Res.* **10**, 4731-4751 (1982).
- Dayhoff, M. O., Barker, W. C. & Hunt, L. T. *Meth. Enzym.* **91**, 524-545 (1983).
- Lockyer, M. J. & Schwartz, R. T. *Molec. biochem. Parasit.* **22**, 101-108 (1987).
- Woo, S. L. C., Lidsky, A. S., Guttler, F., Chandra, T. & Robson, K. J. H. *Nature* **306**, 151-155 (1984).
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294-5299 (1979).
- Robson, K. J. H., Chandra, T., MacGillivray, R. T. A. & Woo, S. L. C. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4701-4705 (1982).

Properdin, the terminal complement components, thrombospondin and the circumsporozoite protein of malaria parasites contain similar sequence motifs

Dimitrios Goundis & Kenneth B. M. Reid

MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

Properdin is a plasma glycoprotein which stabilizes the C3b_nBb enzyme complex of the alternative pathway of the complement system^{1,2}. Unlike the classical pathway, which is initiated by interaction of C1q with the Fc regions of IgG or IgM antibodies in immune complexes, the alternative pathway can be directly activated via binding of C3b to surfaces of foreign organisms^{3,4}. The stabilized C3b_nBbP complex activates components C3 and C5 resulting in opsonization of foreign material (via C3b) and assembly of the membrane attack complex (via C5b) on target cells. Therefore properdin greatly enhances complement-mediated clearance and inactivation mechanisms in both natural and acquired resistance to infection^{1,4}. This paper shows that the primary amino acid sequence of properdin is composed mainly of six repeating motifs, each of ~60 amino acids, and that similar sequences are found in thrombospondin⁵, the circumsporozoite protein of malaria parasites⁶⁻⁸ and regions of the membrane-attack components of complement⁹. These similarities may provide insight into the mechanisms by which parasites avoid host defences mediated by complement.

The amino acid sequence of residues 5-441 of mouse properdin was derived from a complementary DNA (cDNA) clone isolated from a mouse macrophage cDNA library (Fig. 1). The high degree of conservation of sequence between species is indicated by the similarity (~85%) between the derived mouse properdin sequence and the human¹⁰ and rabbit¹¹ sequences over the N-terminal 38 residues. The majority of the properdin amino acid sequence (between residues 53 and 413) is composed of six repeating motifs (of ~60 amino acids, see Fig. 2), that each contain a framework of highly conserved residues in which six cysteines and three tryptophans feature prominently. Initial sequencing studies on the human gene indicate that each motif is precisely encoded by a separate exon. The C-terminal 30 residues show no similarity to the repeating motifs and structure prediction analysis indicates that these residues would adopt an α -helical configuration. The N- and C-terminal 30-40 residues could be involved in the formation of the cyclic oligomers of properdin (mainly dimers, trimers and tetramers) which, as judged from electron microscopy measurements, may result from N-terminal-C-terminal interaction between the monomers of relative molecular mass 56,000 (M_r 56K) which are estimated to be 260 Å in length and 25 Å in diameter^{12,13}. The far UV circular dichroism spectrum of properdin is unusual, however, in that it does not appear to display characteristics associated with extensive α -helix or β -structure¹³.

The framework sequence for the six repeating motifs found in properdin (Fig. 2) shows ~47% identity to each of the three regions of the human thrombospondin molecule on comparison over the entire 58-61 amino acids of each repeat; slightly higher, or lower, values for the percentage identity are obtained on comparison of the individual properdin motifs with any of the thrombospondin motifs (as illustrated by the Align program scores detailed in the legend for Fig. 2). The precise functional role these motifs might play in either molecule has not been clearly defined. Thrombospondin is a 420K glycoprotein composed of three identical chains of ~138K and is thought to be involved in the mediation of cell-to-cell and cell-to-matrix inter-

Fig. 2 Alignment of the six repeating motifs in mouse properdin which occupy residues 53–413 of the molecule. The positions of the first and last amino acids in each motif are shown in brackets on the left. Regions where there are three or more identical aligned residues are enclosed in boxes. Gaps are inserted, or residues inserted above the line (indicated by ∇) to maximize alignment. In *b–e*, residues in common with the properdin framework are boxed. *a*, The conserved, or framework, sequence associated with the properdin repeating motifs. *b*, Comparison of the properdin motif framework with one (residues 420–476) of the three similar regions found in human thrombospondin⁵. *c*, Comparison of the properdin motif framework with one example of the region found at the C-terminal end of the circumsporozoite protein of various malaria parasites (the sequence for *P. falciparum*⁷ is shown). *d*, Comparison of the properdin motif framework with the thrombospondin-related anonymous protein (residues 244–291) found in the blood stage of *P. falciparum*²⁰. *e*, Comparison of the properdin motif framework with one region of C7 (ref 9, residues 8–63) as an example of a typical terminal complement component motif found at the N- and C-terminal ends of C7, C8 α , C8 β and the N-terminal end of C9 (ref. 9).

Methods. The properdin motifs (60 amino acids) were compared by the Align program²⁴ with each other as well as with each of the three motifs of thrombospondin and the two in C7. Scores were based on 100 random runs, bias +6 to the Mutation Data Matrix and a gap penalty of 6. Highly significant alignment scores of 5.0 to 14.0 were obtained between any of the properdin and thrombospondin motifs while lower scores, ranging from 2.0 to 5.0, were obtained between properdin and C7 motifs. The first 26 residues of the properdin motifs, the corresponding regions of those in thrombospondin and the various circumsporozoite proteins were analysed similarly. Significant scores between 3 and 6 were obtained.

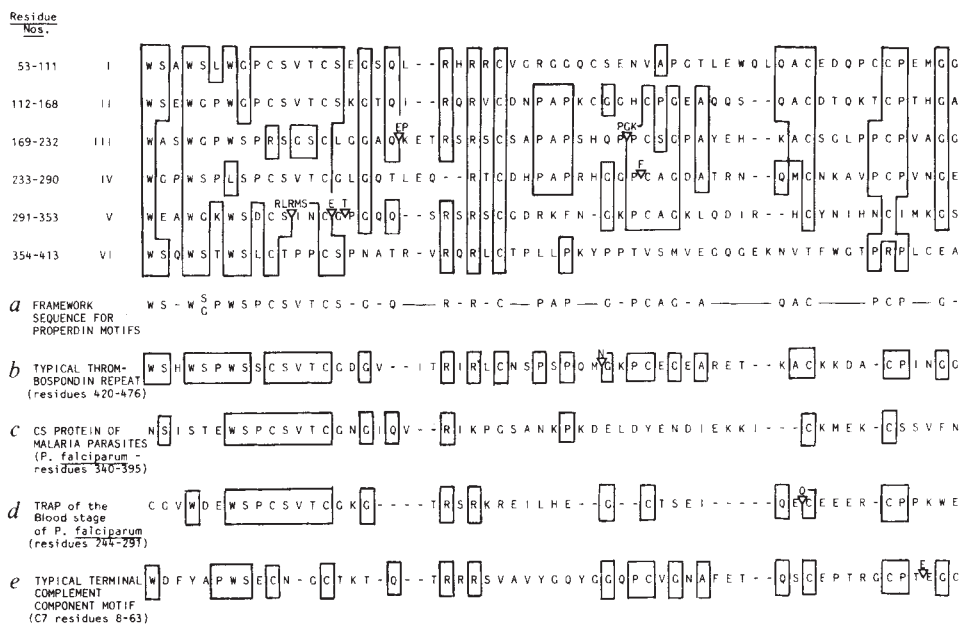
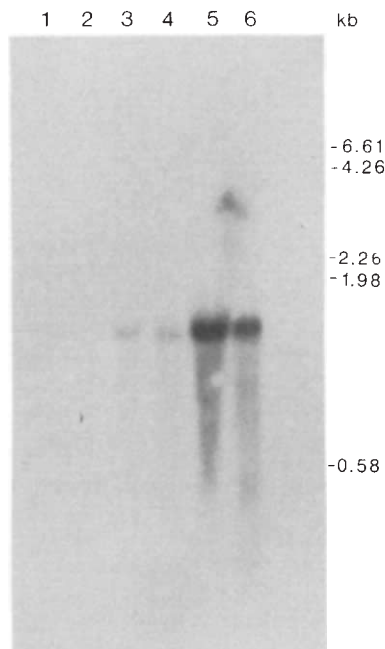


Fig. 3 Northern blot analysis of mouse total RNA from liver (lane 1, 50 μ g; lane 2, 10 μ g), lung (lane 3, 50 μ g), spleen (lane 4, 10 μ g; lane 5, 50 μ g) and activated macrophages (lane 6, 50 μ g). **Methods.** The RNA was isolated by the guanidinium thiocyanate method²⁵ from the blood and tissues of mice which had received an intraperitoneal injection of mineral oil ~48 h previously. A nick-translated 528 bp probe derived from the 5' end of the mouse cDNA clone was used. Marker DNA fragments were derived from a *Hind*III digest of bacteriophage λ DNA and treated in the same fashion as the mRNA samples. Northern blot analysis of the same nitrocellulose filter using a full-length probe for complement component factor B (data not shown) gave strong signals (at 2.6 kb) in the liver lanes (1 and 2), a weak signal in the lung lane (3) and extremely low signals in the spleen lanes (4 and 5) and activated macrophage lane (6), which is consistent with the known primary sites of synthesis of factor B. This showed that the lack of a signal from the liver lanes (1 and 2) using the properdin probes was not due to any defect in amount or quality of RNA loaded in lanes 1 and 2.



between properdin-deficient patients and splenectomized individuals¹⁹ in that both groups suffer from attacks of meningococci, *Haemophilus influenzae* or *Streptococcus pneumoniae* and other encapsulated bacteria^{15–19}. Thus a role of alternative pathway activation and the involvement of properdin is implicated in both cases. The possible role of the properdin/thrombospondin motifs in susceptibility to disease clearly deserves further attention especially in view of the fact that this type of motif is also found in a protein derived from the asexual blood stages

of the malaria parasite²⁰, where it may have some role in host-parasite interaction.

Received 29 April; accepted 24 June 1988.

1. Pillemer, L. *et al. Science* **120**, 279–285 (1954).
2. Fearon, D. T. & Austen, K. F. *J. exp. Med.* **142**, 856–863 (1975).
3. Reid, K. B. M. & Porter, R. R. *A. Rev. Biochem.* **50**, 433–464 (1981).
4. Whaley, K. *Complement in Health and Disease* (MTP Lancaster, 1987).
5. Lawler, J. & Hynes, R. O. *J. cell. Biol.* **103**, 1635–1648 (1986).
6. Ozaki, L. S., Dveck, P., Nussenzweig, R. S., Nussenzweig, V. & Godson, G. N. *Cell* **34**, 815–822 (1983).

7. Dame, J. B. *et al. Science* **225**, 593-599 (1984).
8. Arnot, D. E. *et al. Science* **230**, 815-818 (1985).
9. DiScipio, R. G., Chakravarti, D. N., Müller-Eberhard, H. J. & Fey, G. *J. biol. Chem.* **263**, 549-560 (1988).
10. Reid, K. B. M. & Gagnon, J. *Molec. Immun.* **18**, 949-959 (1981).
11. Nakamo, Y., Matsuda, T., Sakamoto, T. & Motowa, T. *J. immunol. Meth.* **90**, 77-83 (1986).
12. DiScipio, R. G. *Molec. Immun.* **19**, 631-635 (1982).
13. Smith, C. A., Pangburn, M. K., Vogel, C. & Müller-Eberhard, H. J. *J. biol. Chem.* **259**, 4582-4588 (1984).
14. Good, M. F. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 1199-1203 (1988).
15. Sjöholm, A. G., Braconier, J. H. & Sönderström, C. *Clin. exp. Immun.* **50**, 291-297 (1982).
16. Spath, P. J., Misiano, G. & Schaad, U. B. *Complement* **2**, 74 (1985).
17. Densen, P., Weiler, J. M., Griffiths, J. McL. & Hoffmann, L. G. *New Engl. J. Med.* **316**, 922-926 (1987).
18. Gelfand, E. W. *et al. Am. J. Med.* **82**, 671-675 (1987).
19. Bohnsack, J. F. & Brown, E. J. *Am. Rev. Med.* **37**, 49-59 (1986).
20. Robson, K. J. H. *et al. Nature* **335**, 79-82 (1988).
21. Anson, D. S. *et al. EMBO J.* **3**, 1053-1060 (1984).
22. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-569 (1980).
23. Sanger, F., Coulson, A. R., Barend, B. G., Smith, A. J. H. & Roe, B. A. *J. molec. Biol.* **143**, 161-178 (1980).
24. Dayhoff, M. O., Barker, W. C. & Hunt, L. T. *Meth. Enzym.* **91**, 524-545 (1983).
25. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294-5299 (1979).

A PDGF receptor domain essential for mitogenesis but not for many other responses to PDGF

Jaime A. Escobedo & Lewis T. Williams

Department of Medicine, Cardiovascular Research Institute, Howard Hughes Medical Institute, University of California, San Francisco, California, 94143-0724, USA

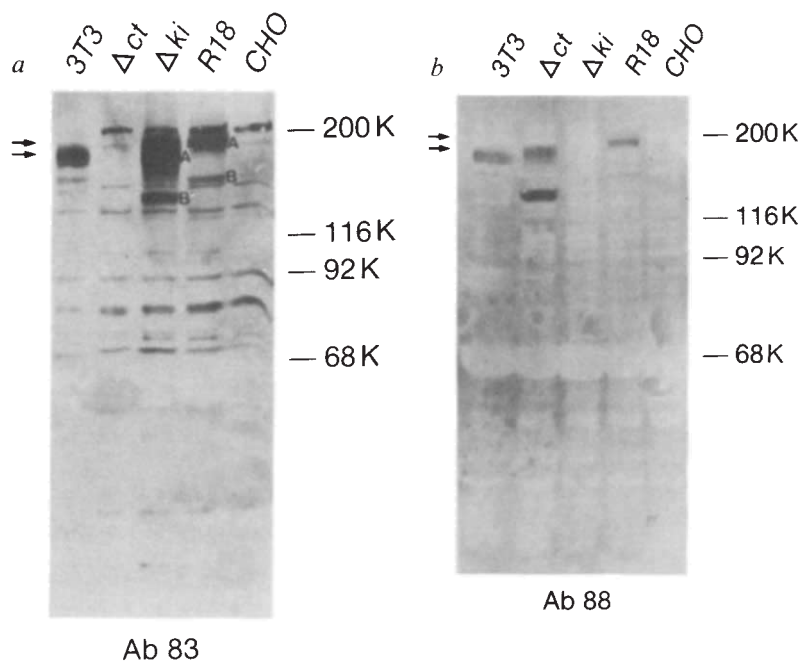
The receptors for mesenchymal growth factors contain tyrosine kinase coding sequences and exhibit ligand-activated tyrosine kinase activity. A variety of mutations of the epidermal growth factor^{1,2}, insulin^{3,4} and platelet-derived growth factor (PDGF) (manuscript in preparation) receptors that have resulted in a loss of tyrosine kinase activity have produced a concomitant loss of growth factor-stimulated DNA synthesis. Comparison of amino

acid sequences of tyrosine kinases shows that these regions in the PDGF receptors in mouse⁵ and human⁶ contain an insert of unknown function. We have deleted this region, and expressed the altered form of the receptor in fibroblasts which lack PDGF receptors. This had no effect on a number of responses to PDGF, but cells bearing the mutant receptor did not proliferate or synthesize DNA in response to PDGF. This demonstrates that the insert is essential in PDGF-induced mitogenesis, and that PDGF stimulation of the mutant receptor tyrosine kinase and phosphatidylinositol turnover are not sufficient to elicit a mitogenic response to PDGF.

One striking aspect of the amino acid sequence of the mouse⁵ and human⁶ PDGF receptors is that the tyrosine kinase coding sequence, recognized by sequence similarity to other known tyrosine kinases, is split into two parts by a 104 amino acid sequence which we have termed the kinase insert (ki) region. We constructed a mutant receptor (Δ ki) in which 82 of the 104 amino acids of the ki region (residues 684-765 using the numbering system of Yarden and colleagues⁵) were deleted. Plasmids containing the mutated receptor complementary DNA (cDNA) sequence under the transcriptional control of the SV40 early promoter were used to express the receptor in Chinese hamster ovary (CHO) cells that normally lack PDGF receptors. Three different stable clones derived from independent transfection experiments were selected for each mutant. As the experimental results using the three clones were identical, data from one representative clone only from each mutant are presented here and are compared with data from a clone that expresses the wild-type receptor (R18) described previously⁷. The level of receptor expression in the Δ ki mutant was higher than the level of receptor expression in cells transfected with the wild-type cDNA (Fig. 1). Immunoblot analysis, using an anti-peptide serum⁸ directed at the region deleted in the Δ ki mutant, confirmed that the expressed protein lacked this determinant and that no wild-type receptor was expressed in the Δ ki transfectants (Fig. 1a, b). The Δ ki mutant cells and CHO cells transfected

Fig. 1 Immunoblot analysis of the kinase insert deletion mutant of the PDGF receptor expressed in CHO cells. Another mutant (Δ ct) in which a sequence of 108 amino acids (922 to 1020) were deleted from the carboxyl terminal region was used as a control. Expression of the Δ ki, Δ ct and mouse wild-type receptors in CHO cells (clone R18) was assessed by immunoblot using antibodies directed against the ki region (Ab83) (a) and carboxyl terminus (ct) (Ab88) (b) of the receptor. CHO parental cells that lack PDGF receptors and Balb/c 3T3 cells expressing native receptors were used as negative and positive control cells respectively. The letters A and A' represent the mature form of the wild-type and the mature form of the Δ ki expressed in CHO cells respectively. Likewise, the precursor forms of the receptor in these cell types have been labelled B and B' respectively. The arrows indicate the native receptor (lower) from Balb/c 3T3 cells and the wild-type receptor (upper) from the transfectant.

Methods. The Δ ki receptor mutant was prepared by digesting a plasmid containing the kinase domain with *SacI* and *BalI* endonucleases. A DNA adaptor was synthesized to restore some deleted amino acids in the kinase region and to link the two kinase domains in frame. The DNA sequence of the mutated receptor was determined by dideoxynucleotide sequencing methods¹². The PDGF mutant receptor cDNA was cloned into the *EcoRI* site of a previously described expression vector⁷ that uses the SV40 early promoter and polyadenylation signals. This plasmid was co-transfected into the CHO cells with a plasmid that contained the gene for neomycin resistance as previously described⁷. Antibiotic resistant clones expressing cell surface mutant or wild-type PDGF receptors were selected by fluorescent activated cell sorting using an antibody directed at the external domain of the PDGF receptor as previously described⁷. Protein lysates from 5×10^5 Δ ki and 5×10^5 wild-type transfected cells were analysed by Western blot analysis using either antibody 83 (Ab83) which recognizes an epitope encoded in the carboxyl terminus (ct), or antibody 88 (Ab88) which recognizes sequences encoded in the kinase insert (ki) region⁸. The immunoblots were developed using a goat anti-rabbit immunoglobulin IgG conjugated with peroxidase and a colour developing system from Biorad. The Δ ct transfectant expresses a PDGF receptor that lacks the epitope for antibody 83 and was used as a control.



Ab 83

Ab 88

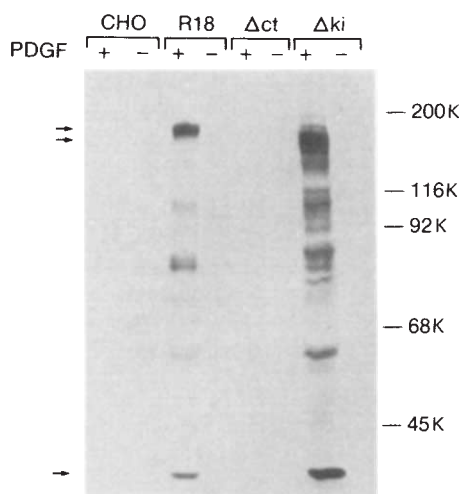


Fig. 2 Receptor tyrosine kinase activity in intact cells. The pattern of PDGF-induced tyrosine phosphorylation in the Δ ki transfectant, wild-type transfectant (R18) and a mutant receptor lacking the carboxyl terminal sequences (Δ ct) was measured by immunoblot using an phosphotyrosine antibody¹⁹. The most prominent bands (see arrows) were a 180–200K receptor protein and a 32K cellular substrate.

Methods. Cells were incubated in the absence and in the presence of 10-nM PDGF. Lysates were prepared as previously described⁷ and were analysed by phosphotyrosine antibody immunoblot.

with the wild-type receptor cDNA, (R18) had high-affinity binding sites for [¹²⁵I] PDGF (approximately 16,000 receptors per Δ ki cell and approximately 5,000 receptors per R18 cell), whereas the parent CHO cells had no specific PDGF binding sites (data not shown). After four hours of PDGF treatment at 37 °C both the Δ ki and wild-type receptor levels were reduced to less than 20% of the level of receptor protein in untreated cells as assessed by immunoblot analysis (data not shown). PDGF-induced receptor down-regulation therefore occurred normally in cells that expressed the mutated receptor.

The ability of the Δ ki mutant to mediate PDGF-stimulated tyrosine phosphorylation of proteins in intact cells was assessed by a phosphotyrosine antibody immunoblot analysis⁹ that measures phosphotyrosine-containing proteins in extracts of PDGF-stimulated cells (Fig. 2). The relative intensity of the signal in extracts of the PDGF-stimulated Δ ki transfectants was similar to the intensity of the signal of the wild-type transfectant (Fig. 2). We showed previously that the phosphotyrosine-containing protein of high relative molecular mass 180,000–200,000 (180–200K) in PDGF-stimulated cells is the receptor itself⁷. As expected the apparent size of the mutant receptor was approximately 10K smaller than the size of the wild-type receptor (Fig. 2). This was caused by the deletion which removed a 82 amino acid segment that included eight tyrosines from the receptor. Whether any of these tyrosines were sites of autophosphorylation of the normal receptor has not been determined. In four separate experiments we found that the actual amount of tyrosine phosphate per receptor was reduced in the Δ ki mutant (20–50% of wild-type levels) even though the total cellular level of receptor tyrosine phosphate was the same or higher in cells expressing the Δ ki mutant than in the wild-type receptor cells. This is consistent with the fact that in most experiments the level of receptor expression in the Δ ki mutant cells was 2–5-fold higher than in the wild-type cells.

In addition to the tyrosine phosphorylated receptor, the phosphotyrosine antibody recognizes an unidentified 32K substrate of PDGF-activated tyrosine kinase. This finding shows that at least one cellular substrate is recognized by both the wild-type and mutant receptor and that the Δ ki mutant receptor is capable of acting as a phosphotransferase *in vivo*. The extent of PDGF-induced phosphorylation of the 32K protein on tyrosine was

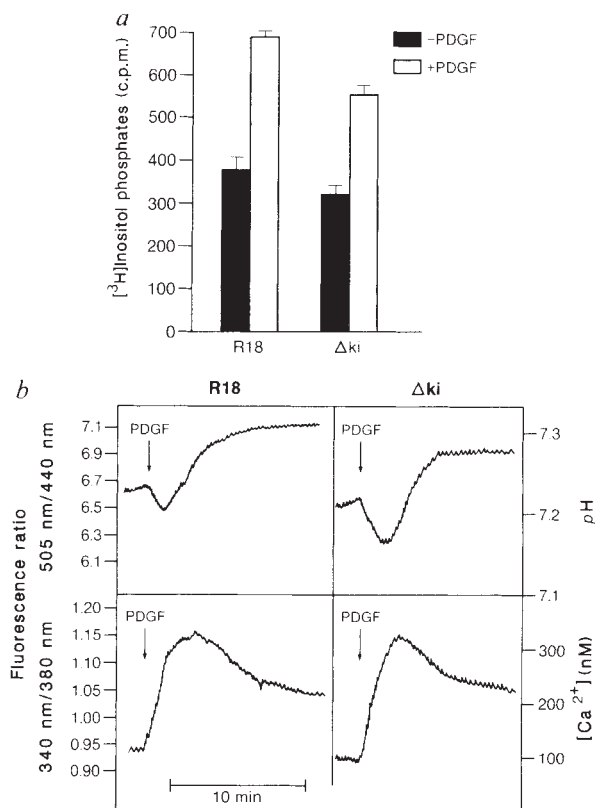


Fig. 3 Rapid responses to PDGF in Δ ki mutant and wild-type (R18) transfectants. **a**, The PDGF-stimulated conversion of phosphoinositides into inositol phosphates, **b**, top, PDGF-induced change in intracellular pH and bottom, PDGF-stimulated intracellular calcium.

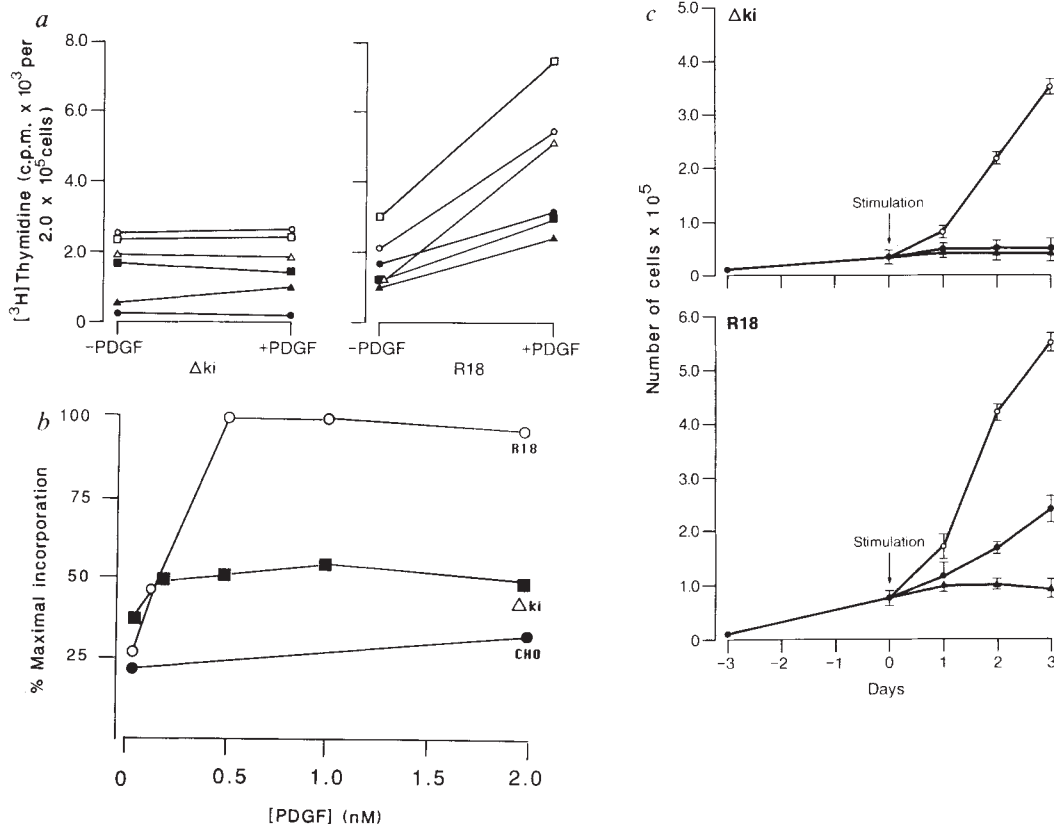
Methods. **a**, Cells were incubated with 20 μ M myo-[³H]-inositol for 24 h. After 30 min stimulation with 10 nM PDGF at 37 °C, the water-soluble inositol phosphates were separated by ion-exchange chromatography using 1 ml columns of AG 108 X8 from Biorad as described previously⁷. Radioactive inositol 1-phosphate, inositol(1,4)diphosphate and inositol(1,4,5)triphosphate were eluted. The total amount of radioactive inositol phosphate from 5×10^5 cells is shown. Each value represents the mean of 3 determinations. **b**, The change in the levels of intracellular pH and calcium in the transfectants after PDGF treatment was determined by measuring the amount of fluorescence emitted by the intracellular probes Fura-2 for calcium and 2,7-bis(carboxyethyl)-5(6)-carboxy-fluorescein, BCECF-AM for pH in a fluorometer SML 8000 as previously described¹¹.

reproducibly (four experiments) higher (5–10-fold on a 'per cell basis') in the Δ ki cells than in wild-type cells, even on a 'per receptor' (approximately 2-fold) basis. In addition a number of other unidentified tyrosine-phosphorylated species were also seen in the PDGF-stimulated Δ ki cells (Fig. 2). The increased PDGF-induced phosphorylation of the 32K protein in the Δ ki cells compared to wild-type receptor cells and the detection of additional unidentified phosphotyrosine-bearing proteins in PDGF-stimulated cells not seen in wild-type cells suggests that the substrate specificity of the PDGF receptor tyrosine kinase is altered in the Δ ki mutant cells. Preliminary data measuring kinase activity using exogenous substrates *in vitro* support this conclusion (W. J. Fantl, J.A.E. and L.T.W., unpublished data).

Binding of PDGF to its receptor stimulates several rapid cellular responses including the hydrolysis of membrane phosphatidylinositol¹⁰, a characteristic alteration of cellular pH (ref. 11) and an increase in intracellular calcium¹¹. The stimulation of these responses by PDGF was similar in the Δ ki and the wild-type receptor transfectants (Fig. 3a, b). Thus the sequences in the ki region are not required for these PDGF-induced responses.

Fig. 4 Proliferative responses of mutant and wild-type transfectants. *a*, Thymidine incorporation into DNA. *b*, Mitogenic dose response curve. *c*, Change in cell number after growth factor stimulation.

Methods. *a*, 2×10^5 cells were incubated with 2 nM PDGF or 10% fetal calf serum. Untreated cells were used as a control. After 18 h of treatment the cells were incubated for 45 min with $1.0 \mu\text{Ci}$ of [^3H]thymidine (25 Ci mmol^{-1} , Dupont, New England Nuclear). Following SDS solubilization the DNA was precipitated with trichloroacetic acid. The pellet was collected by centrifugation and the amount of thymidine incorporated was determined by measuring the radioactivity associated with the DNA pellet¹³. Each value represents the mean of three determinations. The results of 6 independent experiments, each done with and without PDGF, are shown. Fetal calf serum caused an average increase in thymidine incorporation of approximately 4-fold in the Δki mutants and approximately 5-fold in the R18 cells. *b*, 2×10^5 CHO cells, Δki transfectants, or wild-type receptor transfectants (R18) were incubated with increasing amounts of PDGF (0.1, 0.5, 1.0 and 2.0 nM). The amount of [^3H] thymidine incorporated into DNA was measured as above. Each value represent the mean of three independent measurements. Maximal thymidine incorporation stimulated by 10% fetal calf serum was 4-fold over the basal levels. *c*, Cells (2,500) were plated in 24-well plates and grown for three days. After the third day cells were incubated in F-12 Hank's media supplemented with 0.5% bovine serum albumin (Fraction V, Sigma) (triangles), and 10 nM PDGF (solid circles) or 10% fetal calf serum (open circles). Cells were counted in a hemacytometer.



Our original reason for altering PDGF receptor protein sequences in the Δki and other (not shown) mutants was to produce receptor variants that could mediate some, but not all, of the responses to PDGF, and to identify structural domains responsible for specific functions of the receptor. In this context, the result of the analysis of the mitogenic response of the Δki mutant were surprising. Despite the intact rapid PDGF-stimulated responses of the Δki transfectant (PDGF binding, down-regulation, tyrosine autophosphorylation, phosphatidylinositol hydrolysis, pH change, increase in intracellular calcium), these cells do not respond mitogenically to PDGF. In seven independent experiments, PDGF did not stimulate thymidine incorporation into DNA (Fig. 4*a, b*) and did not cause an increase in cell number (Fig. 4*c*) of the Δki transfectant, although serum induced both an increase in DNA synthesis and cell number, presumably through other growth factor receptors that are present in CHO cells (Fig. 4*c* and data not shown). As we recently reported⁷ PDGF stimulated DNA synthesis and an increase in cell number in the wild-type receptor transfectants (Fig. 4*a, b, c*). The relative mitogenic responses of the wild-type and mutant transfectants were consistent over a range of PDGF concentrations (Fig. 4*b*).

These findings were surprising in two respects. First, a large change in the receptor primary sequence—a deletion of 82 amino acids—did not cause a detectable loss of several rapid cellular responses to PDGF. Second, despite the unchanged rapid responses to PDGF, the mitogenic response was abolished by the deletion of the ki domain. This domain may be responsible for mediating the interaction of activated PDGF receptors with other molecules that are important in the mitogenic response to PDGF. Alternatively, this domain may impose a receptor confor-

mation that dictates the substrate specificity of the tyrosine kinase. There may be qualitative alterations of the kinase activity that are not detected by the antiphosphotyrosine immunoblots which only identify the most abundant phosphotyrosine-containing proteins phosphorylated in response to PDGF. A second conclusion from this work is that the stimulation of phosphatidylinositol turnover, pH change and calcium flux change are not sufficient to explain PDGF-induced mitogenesis. Future experiments will focus on the function of this kinase insert domain of the PDGF receptor and of comparable domains of other receptors, such as the colony-stimulating factor-1 receptor.

We thank Dale Milfay and Sutip Navankasatussas for technical assistance, Dr Mark Keating for the antireceptor antibodies, Dr Phil Barr for preparing the DNA adaptor and Dr Harlan Ives for help measuring calcium and pH intracellular levels. We are also grateful to Betty Cheung for her help typing this manuscript. This work was supported by an NIH grant.

Received 28 January; accepted 14 July 1988.

1. Livneh, E. *et al.* *J. biol. Chem.* **261**, 12490–12497 (1986).
2. Chen, W. S. *et al.* *Nature* **328**, 820–823 (1987).
3. Chou, C.-K. *et al.* *J. biol. Chem.* **262**, 1842–1847 (1987).
4. Russell, D. S., Gherzi, R., Johnson, E. L., Chou, C.-K. & Rosen, O. M. *J. biol. Chem.* **262**, 11833–11840 (1987).
5. Yarden, Y. *et al.* *Nature* **323**, 226–232 (1986).
6. Escobedo, J. A. *et al.* *Science* **240**, 1529–1531 (1988).
7. Escobedo, J. A., Keating, M. T., Ives, H. E. & Williams, L. T. *J. biol. Chem.* **263**, 1482–1487 (1988).
8. Keating, M. T. & Williams, L. T. *J. biol. Chem.* **262**, 7932–7937 (1987).
9. Wang, J. Y. *J. molec. cell. Biol.* **5**, 3640–3643 (1985).
10. Berridge, M. J. *Biochem. J.* **212**, 849–858 (1983).
11. Ives, H. E. & Daniel, T. O. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1950–1954 (1987).
12. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
13. Ellis, L. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 5101–5105 (1987).

The proto-oncogene *c-kit* encoding a transmembrane tyrosine kinase receptor maps to the mouse *W* locus

Benoit Chabot*, Dennis A. Stephenson†, Verne M. Chapman†, Peter Besmer‡ & Alan Bernstein*§

* Division of Molecular and Developmental Biology, Mount Sinai Hospital Research Institute, Toronto, Ontario M5G 1X5, Canada

† Department of Molecular and Cellular Biology, Roswell Park Memorial Institute, Buffalo, New York 14263, USA

‡ Laboratory of Molecular Oncology, Memorial Sloan-Kettering Cancer Center, Cornell University Graduate School of Medical Sciences, New York 10021, USA

Mice carrying mutations at the *W* locus located on chromosome 5 are characterized by severe macrocytic anaemia, lack of hair pigmentation and sterility¹. Mutations at this locus appear to affect the proliferation and/or migration of cells during early embryogenesis and result in an intrinsic defect in the haematopoietic stem cell hierarchy^{1,2}. An understanding of the molecular basis of the complex and pleiotropic phenotype in *W* mutant mice would thus provide insights into the important developmental processes of gametogenesis, melanogenesis and haematopoiesis. Here we show that the mouse mutant *W*^{19H} has a deletion of the *c-kit* proto-oncogene. Interspecific backcross analysis demonstrates that the *W* locus is very tightly linked to *c-kit* and that the two loci cannot be segregated at this level of analysis. *c-kit* is the cellular homologue of the oncogene *v-kit* of the HZ4 feline sarcoma virus³ and encodes a transmembrane protein tyrosine kinase receptor that is structurally similar to the receptors for colony-stimulating factor-1 (CSF-1) and platelet derived growth factor^{4,5}. The co-localization of *c-kit* with *W* provides a molecular entry into this important region of the mouse genome. In addition, these observations provide the first example of a germ-line mutation in a mammalian proto-oncogene and implicate the *c-kit* gene as a candidate for the *W* locus.

The work presented here establishes a possible relationship between the *W* and *c-kit* loci and was initiated following the discovery that the *c-kit* proto-oncogene is located on human chromosome 4 at bands 4q11-q21 (refs 4, 6). This region of the human genome encompasses several genes, including *PGM2* (phosphoglucomutase-2), *AFP* (alpha-fetoprotein) and *ALB* (albumin), which are on a conserved region of mouse chromosome 5 that includes the *W* gene locus⁷. To determine whether the *W* locus was linked to the *c-kit* proto-oncogene, we first used a presumed deletion mutant of the *W* locus, *W*^{19H}, that includes *W*, the closely linked Patch (*Ph*) and a recessive lethal (*l*) locus located 1–2 cM (cM = centimorgan, the distance along a chromosome that gives a recombination frequency of 1%) distal to *W* (ref. 8). To investigate whether the *c-kit* gene was deleted in the *W*^{19H} mutant, we crossed *W*^{19H}/+ (101/H × CBA/H) heterozygotes with *Mus spretus* or the laboratory strain C57BL/6J. Restriction fragment length polymorphisms (RFLPs) between these parental stocks were found for the *c-kit* gene using the restriction enzyme *Bgl*II (Fig. 1). F₁ animals from *W*^{19H}/+ and +/+ (*M. spretus* or C57BL/6J) crosses should yield offspring that inherit the *W*^{19H} or the *W*^{19H}/+ allele. *W*^{19H}/+ heterozygotes can be easily distinguished by their characteristic small white patches of hair. Hybridization of genomic DNA with a *v-kit* probe revealed that animals that had inherited the *W*^{19H} allele lacked the *Bgl*II restriction fragment characteristic of the *W*^{19H}/+ parent and exhibited only the *c-kit* *Bgl*II restriction fragment of the *M. spretus* or C57BL/6J

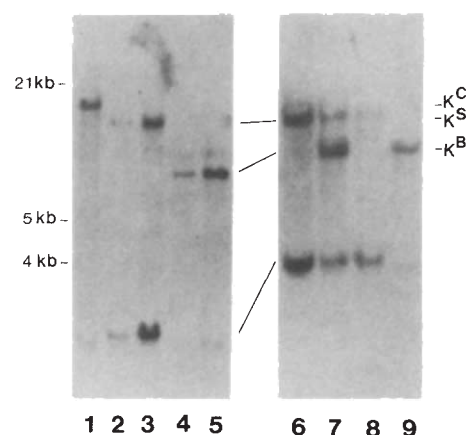


Fig. 1 Exclusive *c-kit* gene contribution from the wild-type allele in various *W*^{19H}/+ strains as detected through RFLP analysis. Our starting stock of *W*^{19H}/+ animals (lane 1) was crossed with *M. spretus* (lane 3) or C57BL/6J (lane 5). The heterozygous F₁ *W*^{19H}/+ × *M. spretus* (lane 2) *W*^{19H}/+ × C57BL/6J (lanes 4 and 9) were analysed. Alternatively, an F₁ (*W*^{19H}/+ × C57BL/6J) animal represented in lane 9 was crossed with *M. spretus* (lane 6), and a *W*^{19H}/+ (lane 8) and a +/+ littermate (lane 7) were analysed. Following DNA extraction from tails, 5 µg of each sample was restricted with *Bgl*II, run on 0.8% agarose gel and blotted on Gene Screen Plus. Blots were hybridized at 42 °C with the feline *v-kit*³ probe labelled by nick-translation (0.3 × 10⁹ c.p.m. µg⁻¹ and washed in 2 × SSC, 1% SDS at 65 °C before autoradiography. K^C, K^S and K^B are the *c-kit* fragments specific to CBA/H, *M. spretus* and C57BL/6J mice, respectively.

parent (Fig. 1). By contrast, Southern blot analysis of genomic DNA from F₁ animals that were phenotypically normal (+/+) gave *c-kit* bands derived from both parents (Fig. 1). These results provide direct molecular evidence that the *W*^{19H} mutation is a deletion that includes the *c-kit* gene.

To define further the linkage between *W* and *c-kit*, F₁ animals obtained from crosses between sash (*W*^{sh}/*W*^{sh}) homozygotes and either *M. domesticus* (inbred strain PAC) or *M. musculus* (inbred strain PWK) were backcrossed to the laboratory strain C57BL/6J congenic carrying the beige mutation on chromosome 13 (Fig. 2a). Although sash homozygotes have only traces of coat pigmentation around the eyes and on the ears, sash heterozygotes have a distinctive white band across the loins providing an unambiguous phenotype to score⁹. The *c-kit* allele contributed by both the PAC and PWK parents could again be distinguished from the laboratory-derived allele by Southern analysis of genomic DNA digested with *Bgl*II (Fig. 2b). Approximately 50% of the backcross progeny were homozygous for the *c-kit* gene whereas the remaining half were *c-kit* heterozygotes, possessing both the wild-derived plus the C57BL/6J banding pattern (Table 1). Thus the *c-kit* fragment variations segregated with the frequency expected of a single autosomal gene in the backcross progeny, indicating that the pattern identified by the *v-kit* probe defined different alleles at the same locus in both the wild- and laboratory-derived parents.

The backcross progeny were then analysed for the segregation of three markers on mouse chromosome 5: *W*, *Kit* and *Afp*. No

Table 1 Segregation of *W* and *c-kit* alleles

Parental class 1	<i>W</i> ^{sh} <i>Kit</i> ^B / <i>W</i> ^{sh} <i>Kit</i> ^B	108
Parental class 2	<i>W</i> ^{sh} <i>Kit</i> ^P / <i>W</i> ^{sh} <i>Kit</i> ^B	112
Recombinant class 1	<i>W</i> ^{sh} <i>Kit</i> ^P / <i>W</i> ^{sh} <i>Kit</i> ^B	0
Recombinant class 2	<i>W</i> ^{sh} <i>Kit</i> ^B / <i>W</i> ^{sh} <i>Kit</i> ^B	0
Total = 220		

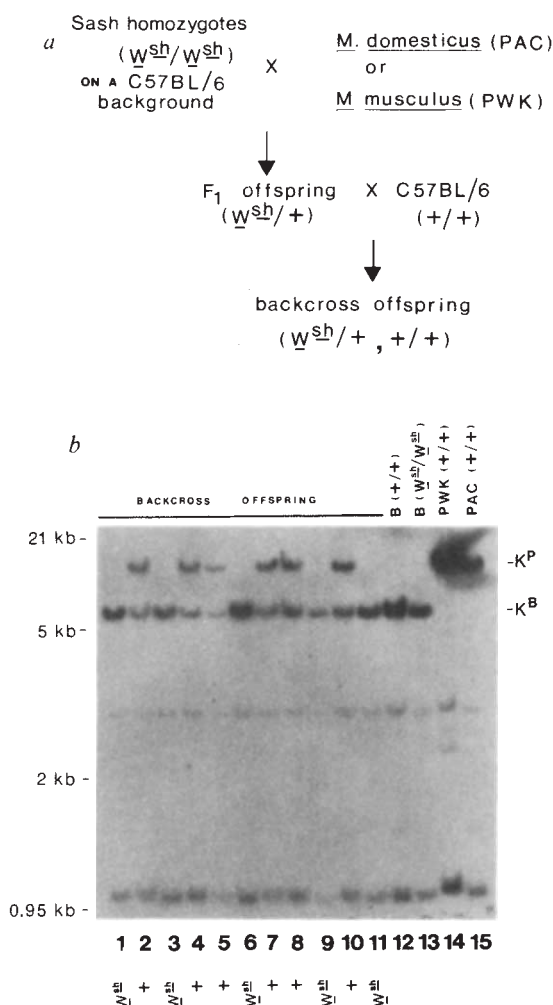


Fig. 2 *a*, Outline of the backcross strategy between W^{sh} homozygous (C57BL/6 background) and either a wild-derived *M. domesticus* (PAC) or an inbred wild-derived *M. musculus* (PWK) strain. PAC and PWK backcross offspring (110 animals each) were analysed independently for the segregation of the *c-kit* gene RFLP and the segregation of the coat-colour marker characteristic of the W^{sh} allele. *b*, An example of the segregation of the RFLPs for the *c-kit* proto-oncogene. DNA was extracted from kidneys, cut with *Bgl*III and analysed as in Fig. 1. Lanes 1–11 represent the last 11 animals of the PWK backcross offspring analysed. Lanes 12–15 correspond to DNA from the parents used in the crosses, B indicates C57BL/6 background. Phenotypic identification of the backcross animals are indicated below the lane numbers. K^P and K^B are *c-kit* fragments specific to PWK/PAC and C57BL/6, respectively.

recombination between *W* and *Kit* was observed in the 220 backcross animals (Table 2 and Fig. 2*b*). In contrast, recombination between *W* or *Kit* and *Afp* was observed in the same 4 animals (Table 2). From the frequency of recombination (4/74), the distance between *W* or *Kit* and *Afp* is 5.4 ± 2.6 cM, a distance consistent with expectations based upon indirect estimates of the *W*–*AFP* linkage¹⁰. Thus, *c-kit* maps with the *W* locus on mouse chromosome 5 between 0 and 1.4 cM at the 95% confidence interval.

The *W* locus and the *c-kit* proto-oncogene are therefore tightly linked. The co-localization of these loci provides an important molecular entry into this region of mouse chromosome 5 that includes not only *W*, but two other very closely linked loci, Patch (*Ph*) and Rump-white (*Rw*) (refs 1, 11), that may also be involved in development and cellular differentiation. These

Table 2 Recombination fractions between *W*, *Kit* and *Afp* on mouse chromosome 5

Locus	<i>W</i>	<i>Kit</i>	<i>Afp</i>
<i>W</i>	—	0/220	4/74
<i>Kit</i>	0/220	—	4/74

Backcross animals were scored for the donor of the *W*, *Kit* and *Afp* genes, based on their coat colour (*W*) and RFLPs generated with *Bgl*III (*Kit* and see Fig. 2*b*) and *Pst*I (*Afp*). Polymorphic bands defined for both the C57BL/6J-derived *Afp* allele and the PWK derived allele were identified using *Pst*I-digested DNA. The C57BL/6J-derived *Afp* allele was defined by 15.5, 0.7 and 0.5 kb bands whereas the PWK allele was defined by *Pst*I bands of size 4.6, 0.7 and 0.5 kb. Thus, the 15.5 and 4.6 kb bands were diagnostic of the two alleles. No polymorphism has been identified that distinguishes the PAC from the C57BL/6J allelic forms.

results also raise the interesting possibility that the *c-kit* gene is an integral part of the *W* locus and is therefore critical to haematopoiesis. Two observations are consistent with this idea. First, the protein product of the *c-kit* gene has striking structural features in common with the CSF-1 receptor (*c-fms*), a known haematopoietic growth factor receptor¹². Second, *c-kit* is known to be expressed in haematopoietic tissues, including bone marrow and spleen⁵.

One might expect that mutations in the ligand for a transmembrane growth factor receptor would result in a phenotype which is almost identical to that of mutations in the receptor itself. Ligand mutations, however, would presumably affect the environment in which stem cells proliferate rather than affecting stem cells directly. Mice bearing mutations at the *Steel* (*Sl*) locus on chromosome 10 are, like *W* mutants, white, sterile and have a severe macrocytic anaemia¹. Bone marrow transplantation experiments have established, however, that the *Steel* defect is mediated through the microenvironment in which stem cells function, whereas the *W* defect is intrinsic to these cells¹³. These contrasting effects of mutation at the *W* and *Sl* loci on haematopoietic stem cells may reflect, therefore, a receptor-ligand relationship between the products of the *W* and *Sl* genes.

The determination of cell fate and the orderly program of development in *Drosophila* involves a number of genes with strong sequence similarity to mammalian proto-oncogenes¹⁴. Our results provide the first example of a germ-line mutation in a mammalian proto-oncogene and suggest that such mutations can also lead to developmental abnormalities in mammals.

We thank Mary Lyon and Peter Glenister for providing the W^{19H} mutant mice, Liz Giddens, Bernadette Keitz and Diane Poslinski for technical assistance, Neil Copeland for many helpful discussions and Martin Breitman, Janet Rossant and Louis Siminovitch for critically reviewing the manuscript. This work was supported by grants from the National Cancer Institute of Canada (NCIC) to A.B., the American Cancer Society to P.B., and the NIH to D.A.S., V.M.C. and P.B. B.C. is a Research Fellow of the NCIC.

Received 3 June; accepted 15 July 1988.

- Russell, E. S. *Adv. Genet.* **20**, 357–459 (1979).
- McCulloch, E. A., Siminovitch, L. & Till, J. E. *Science* **144**, 844–846 (1964).
- Besmer, P. *et al. Nature* **320**, 415–421 (1986).
- Yarden, Y. *et al. EMBO J.* **6**, 3341–3351 (1987).
- Qiu, F. *et al. EMBO J.* **7**, 1003–1011 (1988).
- d'Auriol, L., Mattei, M.-G., André, C. & Galibert, F. *Hum. Genet.* **78**, 374–376 (1988).
- Nadeau, J. H. & Taylor, B. A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 814–818 (1984).
- Lyon, M. F., Glenister, P. H., Loutit, J. F., Evans, E. P. & Peters, J. *Genet. Res.* **44**, 161–168 (1984).
- Lyon, M. F. & Glenister, P. H. *Genet. Res.* **39**, 315–322 (1982).
- Davison, M. T. & Roderick, T. H. *Genetic Variants and Strains of the Laboratory Mouse* (ed. Green, M. C.) 283–313 (Fisher, Stuttgart, 1981).
- Searle, A. G. & Truslove, G. M. *Genet. Res.* **15**, 227–235 (1970).
- Sherr, C. J. *et al. Cell* **41**, 665–676 (1985).
- McCulloch, E. A., Siminovitch, L., Till, J. E., Russell, E. A. & Bernstein, S. E. *Blood* **26**, 399–410 (1965).
- Basler, K. & Hafen, E. *Trends Genet.* **4**, 74–79 (1988).

Cloning of bovine GAP and its interaction with oncogenic *ras* p21

Ursula S. Vogel, Richard A. F. Dixon,
Michael D. Schaber, Ronald E. Diehl,
Mark S. Marshall, Edward M. Scolnick,
Irving S. Sigal & Jackson B. Gibbs

Department of Molecular Biology, Merck Sharp and Dohme Research Laboratories, West Point, Pennsylvania 19486, USA

The plasma membrane-bound mammalian *ras* proteins of relative molecular mass 21,000 (*ras* p21) share biochemical and structural properties with other guanine nucleotide-binding regulatory proteins (G-proteins)¹⁻³. Oncogenic *ras* p21 variants result from amino acid substitutions at specific positions that cause p21 to occur predominantly complexed to GTP *in vivo*. Recently, a GTPase activating protein (GAP) with cytosolic activity has been discovered that stimulates the GTPase activity of normal but not of oncogenic *ras* p21 (ref. 4). GAP might be either a negative regulatory agent which acts further upstream in the regulatory pathway or the downstream target of *ras* p21 (refs 3, 5 and 6). We have identified a protein from bovine brain with apparent relative molecular mass 125,000 that has GAP activity⁷. Here, using pure GAP in a kinetic competition assay, we show that GAP interacts preferentially with the active GTP complexes of both normal and oncogenic Harvey (Ha) *ras* p21 compared with the inactive GDP complexes. We also report the cloning and sequencing of the complementary DNA for bovine GAP. Regions of GAP share amino acid similarity with the noncatalytic domain of adenylate cyclase from the yeast *Saccharomyces cerevisiae*⁸⁻¹⁰ and with regions conserved between phospholipase C-148, the *crk* oncogene product and the nonreceptor tyrosine kinases^{26,27}.

We purified GAP to near-homogeneity from the soluble fraction of bovine brain as previously described⁷. GAP is a monomeric polypeptide with an apparent M_r of 125,000 (125K) and lacking intrinsic GTPase activity; it acts catalytically to stimulate normal Ha p21 GTP hydrolytic activity. In the presence of 2.5 nM GAP, the rate of the conversion of p21-GTP to p21-GDP does not show saturation with concentrations of Ha p21-GTP up to 10 μ M (unpublished observations). To investigate whether GAP could be saturated at higher levels of *ras* p21, we have assayed the ability of purified *ras* p21, bound either to unlabelled GTP or to unlabelled GDP, to compete in assays containing Ha p21-[γ -³²P]GTP and purified bovine brain GAP. As shown in Fig. 1a for normal Ha p21, the GTP complex of Ha p21 competes in the assay with a 50% inhibitory concentration (IC_{50}) of 110 μ M. The GDP complex of Ha p21 was ineffective up to 1.5 mM. Free nucleotide accounted for less than 1% of the total, and the dissociation of bound nucleotide from p21 under the assay conditions (4 minutes at 24 °C in the presence of Mg^{2+}) was negligible. Addition of free GTP or GDP up to 500 μ M inhibited the assay by less than 15%. Inclusion of 50 mM NaCl reduced the competition observed with 100 μ M Ha p21-GTP. The salt effects observed on GAP-stimulated *ras* p21 GTPase activity can be explained by the inhibition of GAP/p21 interaction⁷.

The ability of GAP to interact with various *ras* mutants is critical to the question of whether GAP is an upstream regulator or a downstream target of *ras* p21. Figure 1b compares the abilities to compete in the GAP reaction of the GTP and GDP complexes of the oncogenic protein substituted at position 12 with a valine residue ([Val¹²]Ha p21). The GTP complex competed with an IC_{50} that was 3-fold greater than that for normal *ras* p21. The GDP complex of [Val¹²]Ha p21 had a significantly reduced effect, with 20% inhibition observed at 1.5 mM. A more effective competition was observed with GTP complexes of the oncogenic variants [Arg¹²]Ha p21 and [Leu⁶¹]Ha p21, with IC_{50}

Table 1 Biochemical activity of *E. coli*-expressed GAP

Addition	GAP immunoreactivity	<i>ras</i> p21-GTP hydrolysis (%) Ha	[Val ¹²]Ha
None		<3	<3
Pure bovine brain GAP	+	94	<3
<i>E. coli</i> , pCON	-	4	<3
<i>E. coli</i> , pGAP	+	93	<3

Ha p21-GTP and [Val¹²]Ha p21-GTP were assayed for GAP-stimulated GTP hydrolysis by the nitrocellulose filtration binding method⁷ using p21-[γ -³²P]GTP complex (1,500 Ci per mmol) and purified bovine brain GAP (40 ng) or lysates (48 μ g protein) prepared from *E. coli* cells transformed with the indicated plasmids. Results are expressed as the percentage of total p21-GTP in the assay converted to p21-GDP. Assays were carried out until nearly complete (100% hydrolysis corresponds to 200,000 c.p.m.) by incubation at 24 °C for 5 min. (+) designates positive immunoreactivity by immunoblot analysis. Plasmid pGAP contained the cDNA insert oriented with the region encoding GAP amino acids 128-1,044 in-frame with the β -galactosidase gene of pUC13. pCON contains the cDNA insert in the reverse orientation relative to β -galactosidase. Plasmid-bearing cells were grown at 37 °C for 15 hour in Luria-Bertani medium containing 50 μ g ml⁻¹ ampicillin and 100 μ M isopropyl thiogalactoside. Cell lysis was performed by sequential incubations at 4 °C with 200 μ g ml⁻¹ lysozyme, 0.01% lubrol-PX, and 3 μ g ml⁻¹ DNase I/5 mM MgCl₂ in buffer containing 50 mM Tris-Cl, pH 7.5, 1 mM EDTA, 1 mM dithiothreitol. A supernatant fraction was prepared by centrifugation at 7,500 g for 10 min. Samples were subjected to SDS-PAGE on 7.5% polyacrylamide gels and then electrophoretically transferred onto nitrocellulose. Rabbit polyclonal antiserum was generated using the synthetic peptide Ac-Q-L-P-T-L-G-A-G-L-G-T-V-D-E-K-COOH corresponding to GAP amino acids 139-152 (Fig. 2). The peptide was conjugated via the ϵ -amine of lysine to bovine thyroglobulin (Sigma) at a peptide: thyroglobulin ratio of 1:1 (wt/wt) using 1.25% glutaraldehyde. The coupled peptide (200 μ g equivalent of peptide) in Freund's complete adjuvant was injected into New Zealand white rabbits, and serum was collected after 3 weeks. Immunoblots were probed with anti-GAP peptide serum or pre-immune serum (1:500 dilution) and ¹²⁵I labelled anti-rabbit immunoglobulin (Amersham).

values of 120 μ M and 2 μ M, respectively. The ability of the GTP complexes of these oncogenic variants to compete in a GTP-dependent manner shows that the absence of an increase in their GTPase activities in response to GAP is due to their impaired intrinsic GTP hydrolytic activities, and not because they are unable to interact physically with GAP.

The [Pro¹²]Ha p21 mutant has a high intrinsic GTPase activity that is not affected by GAP⁷ and a greater transforming ability than normal *ras* p21 (ref. 11), and is found to compete for GAP binding in a GTP-dependent manner (60% inhibition at 200 μ M). In contrast, the effector mutants [Asn³³]Ha p21 (ref. 10), [Val¹², Asn³³]Ha p21 (unpublished results), and [Ala³⁸]Ha p21 (refs 5 and 12) did not compete for GAP when their GTP complexes were tested at 0.5 mM.

The abilities of the different *ras* p21 mutants to bind to GAP are consistent with, but not proof of, a function for GAP as the target of *ras* p21. The biological activity of *ras* p21 variants would be determined by the amount of the GTP complex present *in vivo* and the ability of the GTP complex to interact productively with GAP. The amount of the GTP complex present *in vivo* would itself be determined by both the intrinsic and GAP-induced GTPase activities. The roles of these factors are apparent in the following examples. Normal *ras* p21 has a high intrinsic GTPase activity which is accelerated by GAP, binds to GAP as a GTP complex and is mildly transforming. [Pro¹²]Ha p21 has a 2-fold higher intrinsic GTPase activity that is not increased by GAP (overall greater amount of GTP complex), binds to GAP and is more transforming than normal *ras* p21. [Val¹²]Ha p21 has an impaired GTPase activity which is not increased by GAP, binds to GAP and is highly transforming.

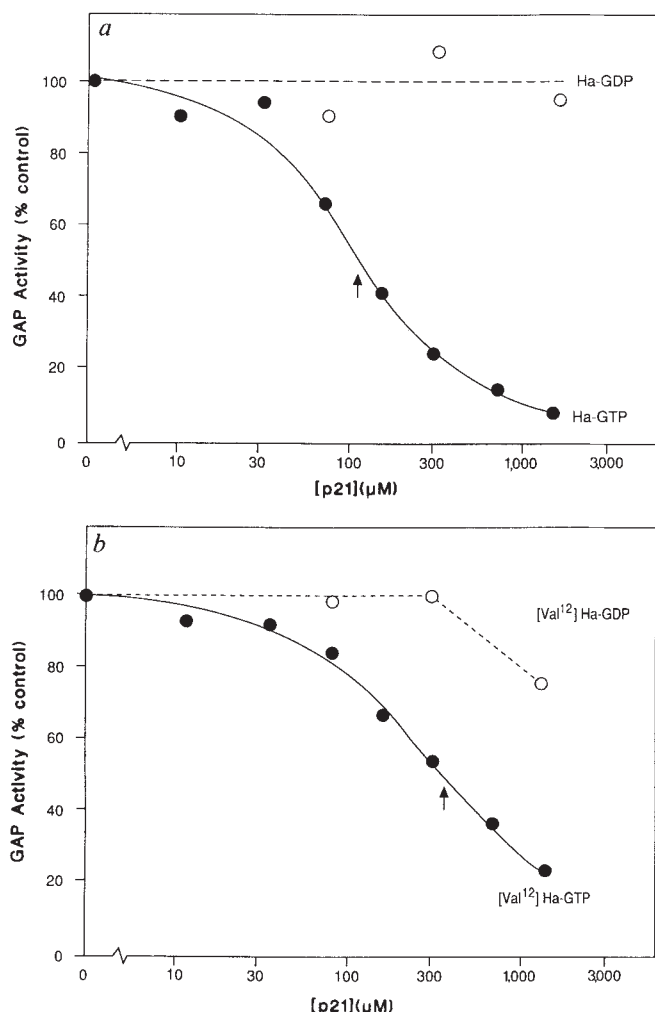


Fig. 1 Interaction between GAP and *ras* p21. GAP-stimulated Ha p21- $[\gamma\text{-}^{32}\text{P}]\text{GTP}$ hydrolysis was assayed in the absence or presence of increasing concentrations of *a*, normal Ha or *b*, oncogenic $[\text{Val}^{12}]\text{Ha}$ p21, complexed to unlabelled GTP or GDP as shown. Results are expressed as a percentage of the control activity (40–60 fmol ^{32}P -labelled orthophosphate produced) in the absence of competing *ras* p21. Arrows denote the concentrations where 50% activity was observed (IC_{50}).

Methods. Non-fusion *ras* p21s were expressed in *E. coli*²², purified²³ in buffer containing 50 mM Tris-Cl, pH 7.5/1 mM EDTA/1 mM dithiothreitol, and then equilibrated with GTP or GDP at 30 °C for 15 min in reactions containing 50 mM Tris-Cl, pH 7.5/1 mM EDTA/1 mM dithiothreitol/60 μM p21/1.5 mM guanine nucleotide. The reaction volumes were repeatedly concentrated in a Centricon-10 (Amicon) by centrifugation at 5,000g and diluted fivefold with 20 mM Na-HEPES, pH 7.5/1 mM MgCl_2 to give a final p21 concentration of 2–4 mM. Final protein concentrations of all reactions were determined by Bio-Rad protein assay and verified by SDS-PAGE analysis with Coomassie Blue staining. The efficiency of the nucleotide-exchange reactions was determined by chromatography on poly(ethylene)imine cellulose containing an ultraviolet indicator dye (Makery-Nagel) with 1 M KH_2PO_4 after heat-denaturation of the *ras* protein 1 M KH_2PO_4 , pH 4. In the absence of heat-denaturation, the bound nucleotide did not migrate from the origin, whereas free nucleotide comigrated with authentic standards. By this analysis, the nucleotide exchange reactions were at least 95% efficient and free nucleotide was <1%. Purified Ha p21 complexed to $[\gamma\text{-}^{32}\text{P}]\text{GTP}$ (ICN Pharmaceuticals) was chromatographed on two sequential PD-10 columns (Pharmacia LKB) to remove free guanine nucleotide. Competition reactions (30 μl) contained 20 mM Na-HEPES, pH 7.5/1 mM MgCl_2 /10 ng purified GAP/8 nM Ha p21- $[\gamma\text{-}^{32}\text{P}]\text{GTP}$ (1,500 Ci per mmol) and the indicated concentration of competing *ras* p21. Solutions were prewarmed to 24 °C for 5 min and reactions were initiated by addition of GAP. After 4 min, the reactions were quenched in an ice-bath by adding 250-μl ice-cold 5% activated charcoal in 50 mM NaH_2PO_4 . Free phosphate in the clarified supernatant (180 μl) was quantitated by Cerenkov counting. Conditions of the assay were linear with respect to time and GAP concentration. In the absence of GAP, GTP hydrolysis was not detectable during the time-frame of the assay.

$[\text{Val}^{12}\text{Asn}^{33}]\text{Ha}$ p21 has a low GTPase activity (as a result of the Val^{12} substitution) which is not increased by GAP, has poor affinity for GAP and is poorly transforming.

To learn more about the structure of GAP, the cDNA encoding the bovine GAP was cloned. Proteolytic cleavage of the purified bovine GAP with trypsin and *Staphylococcus aureus* V-8 protease yielded several peptide fragments whose amino acid sequences were determined (Fig. 2). Oligonucleotides complementary to the GAP messenger RNA sequence predicted from the peptide sequence data were synthesized and used to screen two unamplified bovine brain cDNA libraries (1×10^5 and 5×10^5 recombinants, respectively). Of the 24 hybridization positive clones, all had overlapping restriction maps and several contained a poly(A)⁺ tail at one end. Sequence analysis of 3 overlapping clones revealed an open reading frame of 3,135 base pairs (bp), as defined by 2 adjacent in-frame initiation codons and a 3' termination codon (Fig. 2). This open reading frame would encode a polypeptide with a predicted M_r of 115,742 and an amino acid composition that is almost identical to the experimentally determined amino acid composition of the purified GAP. The potential initiation codon is flanked by sequences that fulfill Kozak's rules¹³ and is preceded by a GC-rich sequence containing several in-frame stop codons. The sequences of all the peptide fragments derived from the purified GAP are present in the predicted protein sequence.

The cDNA insert from one clone, λ GAP1, was subcloned into pUC13 and GAP-related protein was expressed in the *Escherichia coli* strain HB101. The construction encodes a fusion protein of predicted M_r of 107K consisting of the N-terminal 23 amino acids from β-galactosidase fused to amino acids

128–1,044 of GAP. A protein of ~110K was detected by immunoblot analysis of the soluble fraction of extracts from *E. coli* cells transformed with the GAP expression plasmid, but not from control cells, using antisera raised against a peptide corresponding to residues 139–152 from the GAP sequence (see Table 1). Biochemical analysis of the *E. coli* lysates indicated the presence of GAP activity in those samples that expressed GAP protein (see Table 1). Apparently the first 127 residues of GAP are not required for stimulation of p21 GTPase activity.

Previous analyses of GAP activity have shown that GAP is present in most tissues, with the highest concentrations in testes and brain⁷. Immunoblot analysis of tissues using the GAP-specific anti-peptide serum gave a similar result and demonstrated that the size and antigenicity of the GAP isolated from brain was typical of GAP present in other tissues (data not shown). In agreement with these results, blot hybridizations of RNA derived from bovine brain, hamster brain, rat C6 glioma cells and mouse NIH-3T3 cells identified a single 6.0 kilobases (kb) transcript (data not shown), indicating that a similar GAP mRNA is expressed in all of these cells. Blot hybridization of genomic DNA from a variety of species suggests that GAP is probably present as a single copy gene (data not shown).

The predicted structure of GAP has several striking characteristics. The N-terminal 160 residues are extremely hydrophobic and constitute a potential membrane-associated domain. Of the first 130 residues of this domain, 55% are glycine and alanine residues. Immediately C-terminal of this domain is a high concentration of proline residues. In the region amino acids 513–683, proline residues occur at 21–24 residue intervals, possibly reflecting a repeating structural element. The similarity between

Fig. 2 Nucleotide and predicted amino acid sequence of bovine GAP. The nucleotide sequence surrounding the open reading frame contained in several cDNA clones isolated from a bovine brain cDNA library is shown with the first base of the proposed initiation codon as base 1. The complete experimental details for the isolation of the cDNA clones are available from the authors on request. The longest clone isolated contained an additional 483 base pairs 5' to the sequence shown. The 3' untranslated sequence extends 1 kb beyond the sequence shown and ends with a poly(A)⁺ tail. The complete sequence is deposited with GenBank. Numbering of the amino acid sequence begins at the proposed initiator methionine residue. The amino acid sequences for peptides derived from the purified bovine GAP are indicated by underlining, with dashed lines indicating amino acids that were uncertain.

Of the proteins identified in a search of the database for protein sequences similar to GAP using the algorithm defined by Wilbur and Lipman¹⁶, 14 distinct sequences had scores ≥ 50 , representing non-random homology. These sequences were ranked according to the quality of their amino acid alignment as described in the Fig. 3 legend. The sequences with the top five quality scores were the *S. cerevisiae* RNA polymerase, yellow fever virus polypotein, *S. cerevisiae* adenyl cyclase, *Caenorhabditis elegans* myosin heavy chain and complement C3. Of these, the most intriguing is the yeast adenyl cyclase, in the light of evidence supporting its interaction with *ras* proteins in yeast cells⁸⁻¹⁰. There is a 16% identity between the entire GAP

5-aza-2-deoxycytidine, a potent inhibitor of methylation. There is no age difference in the frequency of this reactivation as assayed by HAT⁺ clones, and a more sensitive autoradiographic assay shows only a twofold difference between young and old heterozygotes. Thus, age-related reactivation is not a feature of all X-linked loci, and may have species, tissue and locus-specific determinants.

The age related X-reactivation hypothesis of Wareham *et al.*² was based on studies of female mice heterozygous for a deficiency of ornithine carbamoyl transferase (OCT). Patches of enzyme-positive hepatocytes, identified by histochemical stain, were 50-fold more frequent in mice >1 yr of age than in mice <10 weeks of age. Supporting this are Cattanaach's⁴ observations that female mice, whose coat colour genes that originate on chromosome 7 are silenced by translocation to the inactive X chromosome, gradually acquire extensive pigmentation. A possible mechanism lies in the proposal that cellular ageing is associated with changes in DNA methylation that demethylate inactive loci^{2,5-7} or methylate active ones⁸.

To look for age-effects at a human X-linked locus we chose the HPRT gene, of which only the locus on the active X chromosome is expressed⁹⁻¹¹. As X inactivation occurs at random in each cell, Lesch-Nyhan heterozygotes initially have roughly equal numbers of HPRT⁺ and HPRT⁻ cells (HPRT⁺ allele on active and inactive X respectively). If the HPRT⁺ allele on the inactive chromosome reactivates during ageing, older heterozygotes should have fewer enzyme deficient cells. Any accumulation of mutations affecting HPRT activity is not likely to counteract this effect, as the average mutation rate at this locus, even in older individuals is <10⁻⁵ (ref. 12).

Figure 1 summarizes our analysis of 7,627 cell clones from 41 heterozygotes (2-72 years) in nine Lesch-Nyhan kindreds⁹⁻¹¹. Autoradiography *in situ*¹¹ using the tritiated enzyme substrate ³H-hypoxanthine shows clones (1-2 × 10³ cells each) were either labelled or unlabelled, and were not sectored. The proportion of HPRT⁺ clones is variable at all ages and does not dramatically increase with age (Fig. 1a). Linear regression analysis generates a line with slopes of +0.28 ± 0.16. Although the correlation coefficient is low, ($r = 0.27$, $P = 0.08$), the line intercepts the y axis at 50.9 ± 5.3% and therefore, this population could have been derived from one which at the time that inactivation was initiated, had a random distribution of HPRT⁺ and HPRT⁻ cells. The average number of HPRT⁺ clones in all heterozygotes was 59.0 ± 17.6% (s.d.), significantly different from 50% ($P = 0.002$). This excess of HPRT⁺ clones is observed at all ages, except for the first decade, after which the proportion of HPRT⁺ clones remains relatively constant (Fig. 1b).

The most likely explanation for the excess of HPRT⁺ cells is that they proliferate better than HPRT⁻ cells. This is well documented in haematopoietic tissue from Lesch-Nyhan heterozygotes, in whom HPRT⁻ blood cells are virtually eliminated^{13,14}, except in very young heterozygotes¹⁵. Despite contact feeding between fibroblasts, the lower activity of HPRT in HPRT⁻ cells might confer some proliferative disadvantage.

It is less likely that the excess of HPRT⁺ clones reflects age-related reactivation of the HPRT⁺ allele on the inactive X chromosome because the proportion of HPRT⁺ clones is no greater at age 50-72 yr than at age 20-29 yr. Nevertheless, we looked for age-related reactivation of a silent HPRT⁺ locus by selecting populations of heterozygous cells that were enzyme deficient by growth in 6-thioguanine (6TG). These cells were then grown in non-selective medium for several cell generations before exposure to medium containing hypoxanthine aminopterin thymidine (HAT) which selected those cells that had spontaneously reactivated the silent HPRT⁺ locus and could therefore proliferate in HAT. We similarly tested cells treated with the cytidine analogues, 5-azacytidine (5-azaC) and the more potent methylation inhibitor 5-aza-2 deoxycytidine (5-azaCdr), both known to reactivate the inactive human X in mouse-human hybrids¹⁶ and transformed human cells¹⁷. As methylation locks-in the inactivity of the silent locus¹⁸, we reasoned that age-related

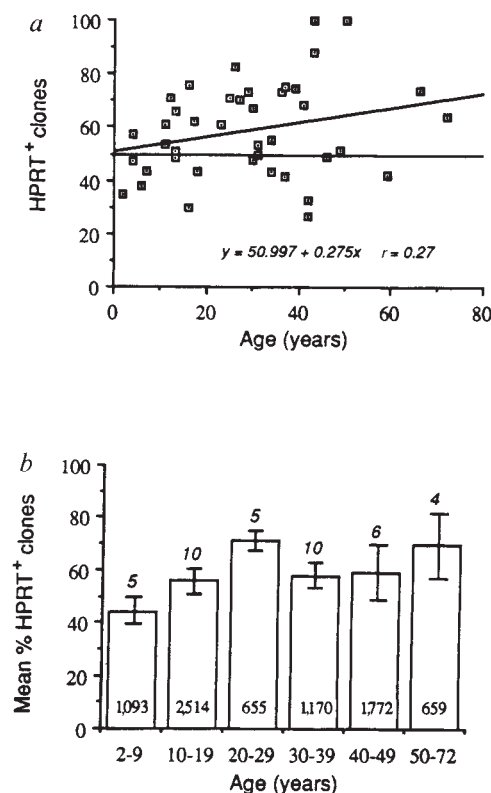


Fig. 1 a, Scattergram showing the proportion of HPRT⁺ clones in skin fibroblast cultures, previously established⁹⁻¹¹ from 41 heterozygotes, plotted according to age. The horizontal line shows the 50% value, expected as a consequence of random X-chromosome inactivation. The equation for the regression line is indicated. b, Bar-graph showing means and standard errors for data grouped according to age of heterozygotes. The number of women is given above the bar, and the number of clones analysed, within the bar.

loss of methylation should facilitate reactivation and this effect would be enhanced due to increased demethylation caused by treatment with the cytidine analogue.

Table 1 shows the results of HAT selection in 19 independent 6TG^r clones from eight heterozygotes (age 8-72 yrs). We obtained neither spontaneous (1.86×10^6 cells tested) nor 5-azaC induced (4.6×10^6 cells tested) HAT⁺ clones in any heterozygote, so the reactivation frequency must be less than 5.4×10^{-7} and 2.2×10^{-7} respectively. On the other hand, we observed eight HAT⁺ clones in cultures treated with 5-azaCdr (2.38×10^7 cells tested). The overall frequency (3.4×10^{-7}) indicates that even 5-azaCdr induced reactivation is rare in heterozygotes. Furthermore, frequencies in old and young women did not differ significantly (2.9×10^{-7} and 4.0×10^{-7} , respectively, $P > 0.1$). The lack of age effect is not due to poorer growth of clones from older women, as cloning efficiencies are similar for HPRT⁺ and HPRT⁻ clones¹¹, and for young and old women (Table 1). The HAT⁺ clones were small (20-200 cells), suggesting that the induced enzyme activity was often insufficient to sustain proliferation in HAT medium. In fact, by labelling enzyme-positive cells *in situ* with ³H-hypoxanthine (Table 2) we detected almost 100-fold more 5-azaCdr induced clumps of labelled cells than HAT⁺ clones (2.3×10^{-5} and 3.4×10^{-7} , respectively) with a twofold difference between older heterozygotes (2.9×10^{-5}) and younger ones (1.3×10^{-5} , $P < 0.005$). Even the more sensitive assay did not detect reactivation in cells not treated with cytidine analogues.

Table 1 Comparison of reactivation frequency in skin fibroblast clones from young and old heterozygotes for HPRT deficiency based on HAT selection

Age (yr)	Clone	Number of cells $\times 10^5$	Treatment	Cloning efficiency*	Number of HAT ^r clones†
8	a	40.0	5azaCdr	11	3 ^{ab}
	b	11.0	5azaCdr	15	0
11	a	2.5	—	25	0
		7.5	5azaC	6	0
		7.0	5azaCdr	3	0
		7.0	5azaCdr	4	0
	b	18.5	5azaCdr	13	0
30	a	10.0	5azaCdr	0.3	1 ^b
34	a	6.0	5azaCdr	10	0
50	a	4.0	—	70	0
		8.5	5azaC	—	0
		6.0	5azaCdr	—	0
	b	2.6	—	57	0
		5.6	5azaC	—	0
		4.5	5azaCdr	—	0
	c	10.0	5azaCdr	11	0
	d	14.5	5azaCdr	21	0
	e	23.0	5azaCdr	24	1 ^c
59	a	1.5	—	28	0
		4.0	5azaC	—	0
		4.5	5azaCdr	—	0
	b	9.5	5azaCdr	—	1 ^b
66	a	1.5	—	74	0
		4.0	5azaC	—	0
		4.0	5azaCdr	—	0
	b	2.5	—	45	0
		7.5	5azaC	29	0
		3.0	5azaCdr	16	0
	c	15.5	5azaCdr	11	1 ^b
		15.0	5azaCdr	9	1 ^a
72	a	4.0	—	65	0
		9.0	5azaC	—	0
		10.0	5azaCdr	—	0
	b	10.0	5azaCdr	4	0
	c	9.0	5azaCdr	4	0

* Cloning efficiency (number of clones obtained/number of cells plated) $\times 100$ was determined from the number of colonies obtained in growth medium in dishes plated with 10–100 cells.

† a, clones with <20 cells; b, clones with >20 , but <100 cells; c, clones with >100 , but <200 cells.

HPRT⁺ clones were obtained by plating 20 heterozygous cells per 60 mm dish, containing 6TG (6×10^{-5} M). After 12–14 days, colonies were picked and transferred to separate dishes in non-selective media. To determine the spontaneous reactivation frequency after pure time in non-selective medium, these 6TG⁺ clonal cell populations were plated into HAT ($0.5\text{--}1 \times 10^5$ cells per 100 mm tissue culture dish), and fixed 20–30 days later. To determine induced frequency, the same clonal populations were plated into growth medium (5×10^5 cells per 100 mm dish), exposed the next day to either 2×10^{-6} M 5-azaC or 2×10^{-7} M 5-azaCdr for 24–48 hr¹⁷ and replated after 3–5 days into HAT-medium (5×10^5 cells per 100 mm dish) before fixing 20–30 days later.

As the only age-effect observed was induced by the demethylating agent, 5-azaCdr, we looked for age-related changes in DNA methylation of the locus in blood leukocytes from normal men and women (Fig. 2). The *Hpa*II methylation pattern for the active gene is well established^{19,20}. Sites in the 5' CpG island and the site(s) at 0.8 kb which affect expression are unmethylated, whereas non-functional *Hpa*II sites 3' to this island are methylated. In the six males (22–81 years) and 14 females (22–89 years) tested, no deviations in this pattern were seen. The methylation patterns for the inactive X were heterogeneous as expected¹⁹ (note multiple 1–3 kb bands in Fig. 2a and individual variation in Fig. 2b), but we saw no age-related differences in band intensity, and no novel bands associated with ageing. This

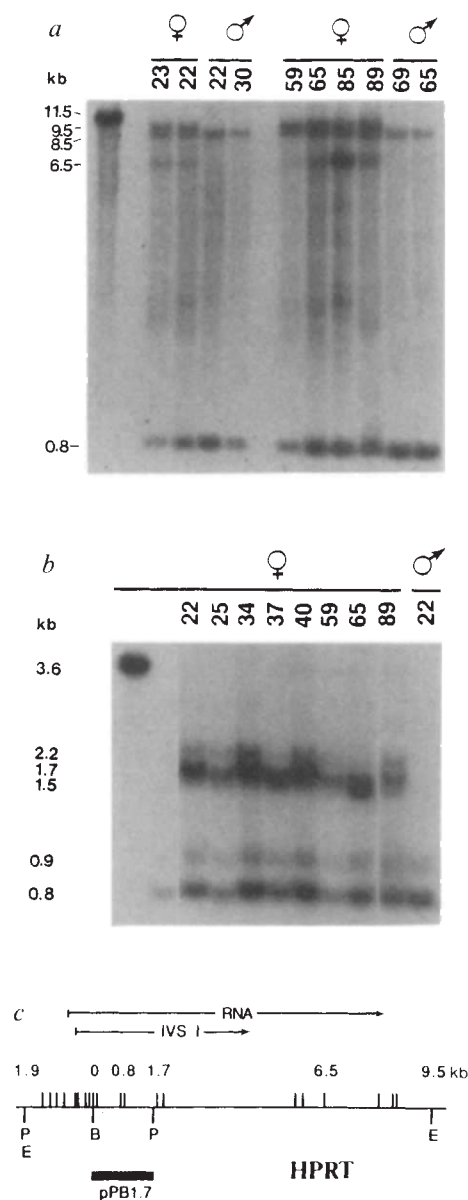


Fig. 2 Autoradiograph of Southern blot showing stability of HPRT methylation. DNA from blood leukocytes of normal men and women, was digested with restriction enzymes and fragments were separated on 0.9% agarose gels¹⁸. Southern blots were hybridized²⁵ with the HPRT intron probe pPB1.7 (ref. 21) labelled with ³²P by the random primer method. The age and sex of the blood donors and chromosomal origin of fragments (Xⁱ—inactive X; X^a—active X), are indicated. *a*, DNA cut with *Eco*RI (lane 1), *Eco*RI+*Msp*I (lane 2) and *Eco*RI+*Hpa*II (all other lanes), and probed with pPB1.7 to examine methylation of 11 kb of the 5' region of HPRT. *b*, DNA cut with *Pst*I (lane 1), *Pst*I+*Msp*I (lane 2) and *Pst*I+*Hpa*II (all other lanes) and probed with pPB1.7 to examine the HPRT 5' CpG cluster. *c* Map (partial) showing relevant *Hpa*II (vertical lines), *Pst*I(P) *Eco*RI(E) and *Bam*HI (B) sites and the probe pPB1.7.

was true not only for the 5' region (Fig. 2b), the only region where demethylation has been shown to be functional¹⁹, but also for the body of the gene (Fig. 2a) that is paradoxically less methylated than the homologous region on the active chromosome. Clearly, methylation of the HPRT locus in blood leukocytes is remarkably stable at all ages.

Our studies of the HPRT locus show that reactivation is

Table 2 Comparison of reactivation frequency in skin fibroblast clones from young and old heterozygotes for HPRT deficiency based on autoradiography

Age (yr)	Clone	Number of cells $\times 10^5$	Treatment	Days in culture	Number of labelled clumps (cells)*
8	a	1.0	—	4	0
		2.0	5azaCdr	27	8 (2-30)
11	b	2.0	5azaCdr	22	0
		2.0	—	1	0
		0.5	5azaC	35	0
	a	2.8	5azaCdr	35	0
		2.0	—	28	0
30	b	3.0	5azaCdr	28	3 (3-10)
		1.5	—	1	0
		2.7	5azaCdr	27	6 (10-20)
34	a	2.0	—	1	0
		1.8	5azaCdr	27	2 (1)
50	a	0.5	5azaC	24	0
		0.5	5azaCdr	24	9 (10-20)
	b	0.38	5azaC	24	0
		0.38	5azaCdr	24	4 (10-20)
	c	2.5	—	1	0
		0.7	5azaCdr	28	0
	d	2.0	—	1	0
		2.5	5azaCdr	22	6 (1-20)
	e	2.0	—	4	0
		2.0	5azaCdr	27	1 (6)
59	a	0.25	5azaC	20	1 (20)
		0.25	5azaCdr	20	2 (10-12)
	b	2.5	—	1	0
		2.4	5azaCdr	28	16 (1-12)
66	a	0.25	5azaC	20	0
		0.25	5azaCdr	20	11 (10-20)
	b	0.5	5azaC	15	0
		0.5	5azaCdr	15	0
	c	4.0	—	1	0
		6.0	5azaCdr	35	9 (1-15)
72	a	0.5	5azaC	20	0
		0.5	5azaCdr	20	0
		2.0	—	1	0
	b	1.6	5azaCdr	28	0
		4.0	5azaCdr	27	4 (1-2)

6TG^r clonal populations, some treated with 5-azaCdr, were maintained either in growth medium for 1-4 days, or in HAT for 15-30 days before labelling. After 16 h incubation in non-selective medium containing ³H-hypoxanthine, cells were fixed in alcohol, subjected to trichloroacetic acid precipitation and autoradiographed¹¹.

* Cells per clump.

extremely rare at any age. While supporting previous evidence that X inactivation in human cells is very stable^{21,22}, our results clearly differ from findings at the OCT locus in mice. Direct comparisons are difficult, as the nature of the evidence is not the same and nor is it known how much OCT activity is detected in the histochemical assay used². Yet, the spontaneous 50-fold increase in patches of cells with OCT activity in ageing mice (>10% of hepatocytes in a low power microscopic field) certainly differs from the induced twofold increase in reactivation of the human HPRT locus. Therefore, it is likely that there is a real difference in reactivation potential of the two loci, perhaps related to species differences. The enhanced reactivation of the

human inactive X chromosome in mouse-human hybrid cells^{16,22} may simply reflect the immortality contributed by cell fusion, or could indicate that inactivation is more stable in human cells. In another species, the opossum *Didelphis virginiana*, we find that the HPRT undergoes spontaneous reactivation with a frequency of 10^{-4} - 10^{-5} HAT^r clones (B.R.M., Jan de Beur, S. & J.A., manuscript in preparation).

Also, inactivation may be more stable in genes like HPRT which are constitutively expressed in all tissues, than in genes like OCT, with limited expression. The differences in methylation that affect expression of housekeeping genes reside in clusters of CpG dinucleotides^{19,23,24}, and these CpG islands are not associated with many tissue-specific genes^{25,26} including OCT. No significant differences in methylation of OCT on active versus inactive X chromosomes in humans²⁷ or mice²⁸ have been reported, making it unlikely that the age-related reexpression of OCT in mice is due to demethylation within the gene. Furthermore, whereas there is evidence that 5-methylcytosine in some cultured cells decreases with age and that 5-azaC-induced demethylation reduces the lifespan of cells^{6,29,30}, there is as yet no compelling evidence that DNA methylation of any specific endogenous single-copy locus changes with age.

Our observations do not support hypotheses that ageing invariably leads to increased rates of spontaneous X-chromosome reactivation or to ubiquitous changes in DNA methylation by documenting the remarkable stability of the human HPRT phenotype at any age, even in cultured cells. The fidelity of HPRT methylation undoubtedly contributes to the stability of inactivation at this locus in human cells at every age.

This work was supported by the NIH. We are grateful to Dr Daniel J. Driscoll for blood specimens from aged individuals and to Dr Peter A. Jones for his gift of 5-aza-2-deoxycytidine.

Received 5 May; accepted 4 July 1988.

1. Lyon, M. F. *Biol. Rev.* **47**, 1-35 (1972).
2. Wareham, K. A., Lyon, M. F., Glenister, P. H. & Williams, E. D. *Nature* **327**, 725-727 (1987).
3. Lesch, M. & Nyhan, W. L. *Am. J. Med.* **36**, 561-570 (1964).
4. Cattanaach, B. M. *Genet. Res.* **23**, 291-306 (1974).
5. Holliday, R. & Tarrant, G. M. *Nature* **238**, 26-30 (1972).
6. Holliday, R. *Expl Cell Res.* **166**, 543-552 (1986).
7. Holliday, R. *Nature* **327**, 661-662 (1987).
8. Harris, M. *Cell* **29**, 483-492 (1982).
9. Migeon, B. R., Der Kaloustian, V. M., Nyhan, W. L., Young, W. J. & Childs, B. *Science* **160**, 425-427 (1968).
10. Migeon, B. R. *Biochem. Genet.* **4**, 377-384 (1970).
11. Migeon, B. R. *Am. J. hum. Genet.* **23**, 199-210 (1971).
12. Trainor, K. J. *et al. in Mechanisms of Ageing and Development* **27**, 83-86 (1984).
13. Nyhan, W. L., Bakay, B., Connor, J. D., Marks, J. F. & Keele, D. K. *Proc. natn. Acad. Sci. U.S.A.* **65**, 214-218 (1970).
14. Dancisi, J. *et al. Life Sci.* **7**, 587-591 (1968).
15. Albertini, R. J. & DeMars, R. *Biochem. Genet.* **11**, 397-411 (1974).
16. Mohandas, T., Sparkes, R. S. & Shapiro, L. J. *Science* **211**, 393-396 (1981).
17. Beggs, A., Axelman, J. & Migeon, B. R. *Somatic Cell Genet.* **12**, 585-594 (1986).
18. Kaslow, D. C. & Migeon, B. R. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6210-6214 (1987).
19. Wolf, S. F., Jolly, D. J., Lunnen, K. D., Friedmann, T. & Migeon, B. R. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2806-2810 (1984).
20. Yen, P. H., Patel, P., Chinault, A. C., Mohandas, T. & Shapiro, L. J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1759-1763 (1984).
21. Migeon, B. R. *Nature* **239**, 87-89 (1972).
22. Wolf, S. F. & Migeon, B. R. *Nature* **295**, 667-671 (1982).
23. Wolf, S. F. *et al. Nucleic Acids Res.* **12**, 9333-9348 (1984).
24. Riggs, A. D., Singer-Sam, J. & Keith, D. H. in *Biology and Biochemistry of DNA Methylation* (eds Cantoni, G. L. & Razin, A.) 211-222 (Liss, New York, 1985).
25. Bird, A. P. *Nature* **321**, 209-213 (1986).
26. Ruta Cullen, C., Hubberman, P., Kaslow, D. C. & Migeon, B. R. *EMBO J.* **9**, 2223-2229 (1986).
27. Hannibal, M. C., Cullen, C. R., Kaslow, D. C., Davies, K. E. & Migeon, B. R. *Am. J. hum. Genet.* **39**, A201 (1986).
28. Mullins, L. J., Veres, G., Caskey, C. T. & Chapman, V. *Molec. cell. Biol.* **7**, 3916-3922 (1987).
29. Wilson, V. L., Smith, R. A., Mas, S. & Cutler, R. G. *J. biol. Chem.* **262**, 9948-9951 (1987).
30. Fairweather, D. S., Fox, M. & Margison, G. P. *Expl Cell Res.* **168**, 153-159 (1987).

Notes on X-ray microanalysis

R. Harrison

The technique of X-ray microanalysis has matured from its humble beginnings to become an essential tool to both research and industry.

THE field of X-ray microanalysis has progressed rapidly since the discovery of X-rays by the German physicist Roentgen in 1895. In the early days X-rays were measured by Geiger counters. Later, wavelength-dispersive spectrometers (WDS) were developed, which analysed X-rays according to wavelength using crystal spectrometers. The revolution in X-ray microanalysis started about 20 years ago with the development of the energy-resolving spectrometer which used a silicon/lithium-drifted solid state detector¹. The speed and simplicity of the energy-dispersive spectrometer (EDS) has made it the most common means of X-ray measurement to be used with electron microscopes.

Despite detection limits ten times lower than EDS (0.01–0.05 per cent for WDS, 0.1–0.5 per cent for EDS) and the lack of peak overlapping problems, wavelength-dispersive systems are less commonly used today because their advantages are offset by the slowness and complexity of the technique involved, and the higher cost of the equipment. An ideal microanalysis system would consist of both energy-dispersive and wavelength-dispersive spectrometers, where the WDS would be used for trace elements and the EDS for major elements.

X-ray microanalysis begins when X-rays generated by the interaction of the electron beam in the microscope and the specimen are collected by a detector. The depth from which X-rays can be generated is dependent upon both the density of the material and the energy of the electron beam (the accelerating voltage of the microscope). The X-rays are produced by the displacement of electrons within the shells of the atoms comprising the sample, which release energy at a level characteristic of each atom's atomic number.

Energy-dispersive

In the EDS system, the silicon/lithium-drifted (Si/Li) detector converts X-ray energy into a pulse of electrical charge using the photoelectric effect. The pulse of charge is converted to a voltage pulse and amplified by a preamplifier and a field effect transistor (FET), before being passed to a processor for further amplification and conversion into a digital signal for measurement and display. The detector is the most sensitive part of the system, and a great deal of research has gone into methods to reduce noise and improve resolution. The detector is cooled by a

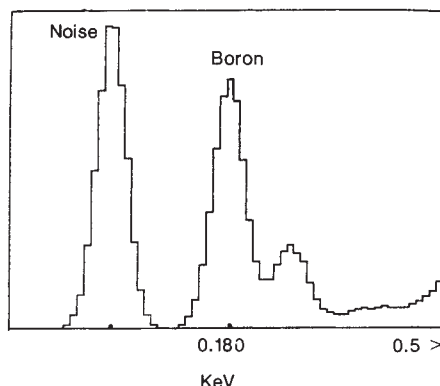


Fig. 1 Boron K_{α} spectrum from a high-purity germanium detector.

liquid nitrogen cryostat to minimize noise and lithium mobility in the crystal. The recent development of more efficient preamplifiers and improved FET designs has further minimized noise and yielded detector resolutions of less than 130 eV.

Because the complete detector assembly must be kept under a vacuum, an entrance window which is transparent to low-energy X-rays but capable of withstanding atmospheric pressure must be present. The normal solution is to use a beryllium foil, usually about 8 μm thick, which allows for 50 per cent transmission of X-rays at 1 keV. The resulting attenuation of X-rays limits the analysis to elements of atomic number 11 (sodium) and above. This limitation has been overcome by the development of detectors which allow the beryllium window to totally be removed or replaced by an ultra-thin window, when the microscope chamber is under vacuum. These light element detectors extend the detector limits down to atomic number 4, beryllium. Use of the ultra-thin window removes the risk of contaminating the Si/Li crystal with volatiles from the microscope and prevents light from entering the detector, which would degrade the resolution. Some attenuation of X-rays caused by the window material is inevitable, but this is only significant when approaching detection limits. A variety of materials are available for ultra-thin windows and the windows can easily be substituted or replaced by an operator.

While the use of Si/Li detectors with scanning electron microscopes (SEM) has become routine, the development of detectors for transmission electron microscopes (TEM) presents an extra set of problems. In order to ensure optimal X-ray collection, high-take-off angle detectors which possess relatively small

probes have been developed.

In an environment where the accelerating voltage may be 100 kV or greater, the Si/Li crystal is susceptible to damage by overloading. To protect the crystal, an automatic shutter mechanism can be interlocked to the magnification control, and to a special circuit which monitors for overload in the detector electronics.

The recent development of detectors that use germanium instead of Si/Li crystals has significantly increased the efficiency range of the instruments, permitting the detection of X-rays of much higher energy with improved resolution². The latest germanium detectors exhibit an efficiency previously unseen at the lower end of the spectrum, down to boron. Although germanium detectors currently show the greatest potential for certain specific applications, it is likely that in the future they will become much more widely used.

Wavelength-dispersive

In the WDS system, X-rays are diffracted by a crystal and detected by a gas-filled proportional counter. As the angle of incidence is varied, different wavelengths are diffracted into the counter, according to the Bragg equation. Spectrometers used today are generally of the linear fully-focusing type: the crystal moves along a linear path while being rotated, and the counter is swung through twice the angle of the crystal. Four crystals of different d -spacings are usually used to cover the full wavelength range.

WDS has an important role in microanalysis because of its improved sensitivity, precision and light element detection. Simultaneous energy-dispersive and wavelength-dispersive detection is now possible through the use of integrated, computer-controlled energy-dispersive/wavelength-dispersive detector systems, which combine the speed of energy-dispersive detectors for major elements with the precision of wavelength-dispersive detectors for minor elements.

Both energy-dispersive detection and wavelength detection systems are controlled by increasingly sophisticated computers, which process incoming data to identify and quantify the elements within a sample. Modern software has been designed to simplify analysis for even the most inexperienced operator: many systems can be used with no real knowledge of computing. The computer systems are capable of controlling the

electron beam, the specimen stage and the spectrometers. Automation procedures are extremely useful for the repetitive analysis of a large number of samples — particularly where energy-dispersive detection and wavelength detection are being performed simultaneously — and leave the operator free for other purposes. The computers are also capable of running more than one task at a time, so the operator can process material whilst new data are being collected.

The main use of X-ray microanalysis systems is in the qualitative and quantitative determination of the composition of sample materials. However, more complex programs for mapping elemental distributions using both energy-dispersive and wavelength-dispersive spectrometers are now commonplace, and recently quantitative elemental mapping with on-line digital filtering³ and drift-corrected mapping⁴ have become possible. Many microanalysis systems are also image analysers, allowing information on features to be combined with elemental information for classification purposes.

The X-ray fluorescence method may be used for inexpensive trace element detection and quantitation when a WDS system is not available⁵. In this technique, a foil is introduced into the path of the electron beam to generate primary X-ray radiation which fluoresces the specimen. The fluoresced X-rays are detected in the usual way by the EDS detector, down to detection limits of 10 p.p.m.

The applications for X-ray microanalysis are innumerable, because the technique is used not only in research, but widely in industry, in all branches of science and technology. The technique is used by metallurgists examining alloys for fatigue, chemists developing catalysts, biologists involved in tissue studies, geologists exploring for mineral and oil deposits, and archaeologists examining artefacts. The quality control of the production and manufacture of textiles, drugs, paints, plastics, semiconductors, and the preparation of food products provides another large market. Forensic science employs the technique to analyse materials ranging from gun-shot residues to glass and paper. With minimal sample preparation, almost any material that can withstand the conditions of the electron microscope can be analysed by this versatile technique. □

Ruth Harrison is at Link Analytical, Halifax Road, High Wycombe, Buckinghamshire HP12 3SE. For more information, fill in reader service number 100.

1. Fitzgerald, R., Keil, K. & Heinrich, K.F.J. *Science* **159**, 528 (1968).
2. Cox, C.E., Lowe, B.J. & Sareen, R.A. *IEEE Trans. Nuclear Sci.* **38**, 28 (1987).
3. Statham, P.J. in *Analytical Electron Microscopy* (ed. Joy, D.C.) 187–190 (San Francisco Press, 1987).
4. Vale, S.H. & Statham, R.A. *Proc. Xth Intl Cong. Electron Microsc.*, 573, Kyoto, 1986.
5. Eckert, R. *Scanning* **8**, 232–238 (1986).

EM at Eurem88

A vast array of electron microscopes and instruments for the preparation of samples for analysis will be on display next week at Eurem88 in York, United Kingdom.

IN stand 24, Oxford Instruments will be showing off its range of accessories for electron microscopy. Sure to be on display is the company's updated range of Hexland **cryo-preparation and cryo-transfer systems** for use with both scanning and transmission electron microscopes (*Reader Service No. 101*). The \$30,500 (US) model CT1000A Cryotrans is a system designed to interface directly to a scanning electron microscope, allowing loading and cryopreparation for freeze fracture, freeze replication, ion beam etching or coating, sputter evaporative coating and electron beam evaporations to be carried out in one unit. The \$24,435 (US) model CT3000A system for transmission electron microscopy employs a cooling holder designed for frost-free transfer of frozen sections or liquids.

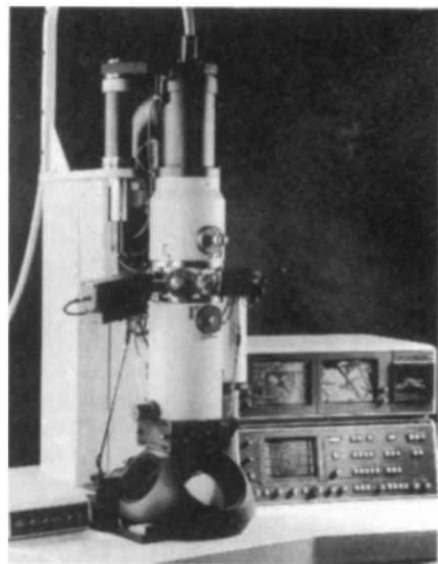
Link Analytical will be displaying its model AN10000 **X-ray microanalysis equipment** on stand 1 (*Reader Service No. 102*). The £25,000–80,000 (UK) AN10000 combines Link's hardware with software running under DEMON, a specially written operating system. Link Analytical says the AN10000 was the first system to incorporate a Winchester hard disk as standard. Current versions use 3.5-inch micro-floppy disk drives; an optional 80 Mbyte Winchester diskpack is available for archiving large amounts of data. The display has a 512 × 512 pixel area for the display of electron microscope images, and electron microscope settings such as beam positioning and stage-motor movement can be performed using the mouse. Software is available for the AN10000 for applications such as bulk specimen analysis, quantitative X-ray microanalysis of thin biological specimens, and the analysis of thin metallurgical specimens.

Electron microscopes

Carl Zeiss, exhibiting in stand 5, has a **computerized scanning electron micro-**

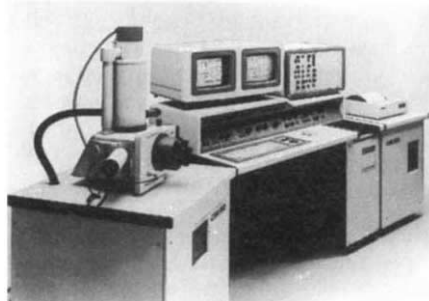
scope for research and industry (*Reader Service No. 103*). The £175,000 (UK) model CSM950 has an integrated image analysis system and program-controlled microscope operation. Zeiss developed a new motorized eucentric six-axis stage for the CSM950, with features such as a 120 mm × 120 mm travelled range and a velocity of 30 mm per second. The computer system allows a maximum of 120 specimen locations to be stored and relocated with a reproducibility of ±0.5 µm, says Zeiss. The image analysis system is comprised of an image processor and a 4 Mbyte framestore for image resolution of 2048 × 2048 pixels, and the separate computer features a hard disk drive and a floppy disk drive. Software is available for grey image processing, arithmetics, segmentation, interactive image operations, binary and multiphase image operations, and the determination of geometrical measuring parameters.

Philips Analytical will be filling stands 4 through 8 with its range of instruments



The CM12 STEM from Philips.

covering scanning, transmission, and scanning transmission electron microscopy. Philips' model CM12 **scanning transmission microscope** combines both electron microscopy techniques in one unit (*Reader Service No. 104*). A key feature of the \$340,000 (US) CM12 is its four-condenser lens system, encompassing spotsizes, intensity, mini-condenser objective, and condenser objective lenses. An interactive softkey microcontroller displays the microscope's status, magnification and other instrument parameters.



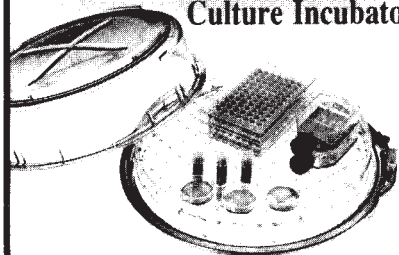
The Zeiss CSM950 SEM.

The mini-condenser lens and the micro-controller allow the microscope to be switched from the three electron beam modes quickly and precisely, says Philips.

Nissei Sangyo America Ltd/Hitachi has announced the introduction of its new model S-2500 scanning electron microscope, on exhibit in stand 3 (*Reader Service*

ADVERTISEMENTS

Economical Modular Tissue Culture Incubator



• **Safe** for AIDS research • **Reliable** for In-Vitro fertilization
• **Versatile** enough for thousands of lab uses
Easy to use – Just flush with gas mixture, seal and place at approp. temp. Perform multiple experiments simultaneously w/out cross contamination. Thousands in use worldwide.
For information contact: **Flow Laboratories, Ltd.** or, **Billups-Rothenberg, Inc.** • P.O. Box 977
Del Mar, CA 92014-0977 USA • (619) 755-3309

Reader Service No.25



WANT THE PERFECT COMBINATION FOR XENOPUS OOCYTE INJECTION? SINGER MK1 MICROMANIPULATOR PLUS INJECT + MATIC

For new brochures write to:
SINGER INSTRUMENTS, TREBOROUGH
LODGE, ROADWATER, WATCHET,
SOMERSET, TA23 0QL.
Tel: (0984) 40226.
Telex: 337492 Comcab G

Reader Service No.50

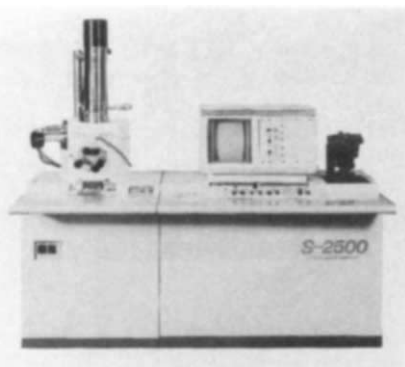
Transfer of DNA/RNA to solid supports, Southern/Northern blots and filter hybridization



Our accurately temperature-controlled rotisserie oven gives **important advantages** as against bag: 2.4-3 ml hybridization solution. Lower background. Higher intensity of bands. Multiple hybridizations possible. Uniform distribution of solution.

**BACHOVER GmbH P.O. Box 7089
D-7410 Reutlingen W-Germany**

Reader Service No.19



Hitachi's model S-2500 SEM.

vice No. 105). The \$89,000 (US) S-2500 has high take-off angle ports for both wavelength-dispersive X-ray analysis and energy-dispersive X-ray analysis. Its specimen chamber will accommodate an 8-inch specimen, and has a stage that will travel up to 100 × 100 mm. The instrument also has a completely automatic x-y stigmator which corrects astigmatism with the touch of a button: together with the auto focus feature, Hitachi says even a novice can take high-resolution photographs. The S-2500 retains the successful features of the older model S-570, including an in-the-lens detection system, menu-driven software and dual secondary detectors.

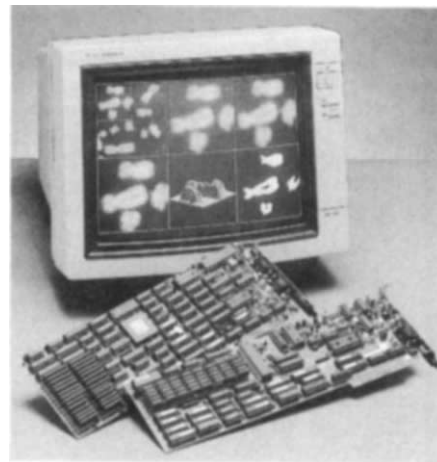
Image processing

Cambridge Instruments has a new **image analysis system** that it says enables the manipulation, processing and measurement of full-resolution grey images — the Quantimet 520 (*Reader Service No. 106*). The \$24,100 (US) Quantimet 520 stores and averages video input images after shading correction, to reduce noise and produce a cleaner image for processing. Point operations work on a single pixel to modify contrast; operations include transformations such as log, exponential, square, inverse and histogram normalization. Digital filtering is provided by Laplace, Sobel, Roberts, and both maximum and minimum transformations. For convolution transformations, users may design their own variable-sized kernels. The system's Microsemper software provides geometric operations such as magnifying, sectioning and rotating the image, the production of power spectra and histograms, and a range of Fourier transformations and correlations. Cambridge Instruments is in stand 7.

Sight Systems will be exhibiting its Oasis 1500 **video archive system** in stand 46 (*Reader Service No. 107*). The £10,000-15,000 (UK) Oasis 1500 system is based on an IBM PC XT or compatible, and captures frames of monochrome video information from a standard video signal for storage on a removable optical disk. Sight Systems says up to 750 different images per disk can be recalled and archived in

seconds. The system can be networked with up to four terminals, making the central information store available to up to four different users. The Oasis 1500 can also be connected to a video printer sold by Sight Systems which can generate finished prints in less than 18 seconds, at less than the cost of conventional instant photography, says the company, making it an economical alternative for laboratories taking pictures from a scanning electron microscope.

Synoptics Ltd is offering two new **framestores** and complementary software designed to enhance the image processing capabilities of IBM PCs and compatibles (*Reader Service No. 108*). Both the £5,500 (UK) Synapse and the £7,500 (UK) Synergy framestores provide real-time television signal imaging, but Synergy also



Two new framestores from Synoptics Ltd.

has slow-scanning capability and its own on-board processor, which Synoptics says is ten times faster than a PC AT or XT. Synergy's image enhancement features include convolution and interpolation, the elimination of noise interference by digital signal filtering through image data averaging and weighting, and zonal or feature colouring. Synoptics is in stand 23.

The Crystal-Sapphire **image processor** will be on display in Quantel Ltd's stand, number 2 (*Reader Service No. 109*). The Crystal-Sapphire system was developed for the enhancement of images obtained through the video output of industrial X-ray systems, optical and transmission microscopes, and other potentially low-grade video sources. The £20,000 (UK) system provides gamma correction for a high degree of contrast, digital shading correction, noise reduction by integration or averaging, and an interactive zoom/roam feature.

These notes are compiled by Carol Ezzell from information provided by the manufacturers. To obtain further details about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.

The school-leaver time bomb

Richard Pearson

The decline in the number of school-leavers means that to remain competitive, employers will have to change their strategies for recruitment and training.

EARLIER this year, one in eight British employers surveyed thought that the number of school-leavers would increase over the next five years, while a further one in four could not give any estimate of what was likely to happen (Fig. 1). Many of them were major recruiters of school-leavers. Few companies are considering adjusting their recruitment policies. Yet it is well known that the decline in numbers of school-leavers will be dramatic, over 35 per cent in the decade to 1993 (*Nature* 329, 470; 1987). The decline is most pronounced in the high unemployment areas of the North and the Midlands, and weakest in the booming South-East. The fall is likely to be highest among the least qualified, again auguring well for the unemployment problem. Even so, the drop of more than 20 per cent among those with five O levels and two A levels is dramatic, particularly as the top 20 recruiters alone take half of those with two A levels going into jobs.

The government is now embarking on a campaign to alert employers to these changes, which are being mirrored in most of the Western world. What then should recruiters be doing about it, and where is the decline likely to hit hardest?

Employers recruit young people for a variety of reasons, for example for their skills, as trainees, because they are cheap

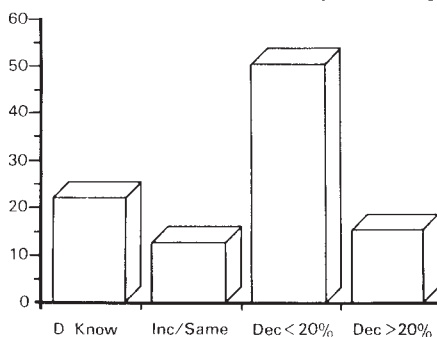


Fig. 1 In answer to the question "What will happen to youth labour supply by 1993?", few employers foresee the large decrease. Data from *Employer Response to the Decline in School Leavers*, IMS, 1988.

or because they are available. The Youth Training Scheme (YTS) is now a key entry channel for young people into many jobs; the direct recruitment of 16–17-year-olds is falling in importance as YTS is increasingly seen as the basis for developing broad-based abilities in young people. Where higher qualification levels are required, 18–21-year-olds are sought. The

same is true in, for example, shift working, where legislation restricts the use of younger workers.

Despite the shortage of school-leavers, the total labour force will grow by nearly a million over the period to 1995, with the growth being dominated by women, many of whom will be returning to work. Yet for most companies, re-entrants to the labour market are not seen as a significant source of recruitment. Internal appointments and retraining also remain a largely untapped source of new entrants, despite the advantages of using staff who are already familiar with many aspects of the organization. Although many employers are clearly dependent on young people, there seems no objective reason for their being the chief and, in some cases, the exclusive source of entrants.

This may be one reason for the general lack of awareness among employers of the marked changes coming to school-leaver supply, and their lack of preparation for the consequences (Fig. 2). Employers in the public sector, financial services, construction and transport were generally the best informed, with retailers, hoteliers and caterers the least well informed. Awareness among the larger companies often did not spread beyond the head office or the personnel function.

Given this lack of awareness, it is not surprising that few companies were planning to change their recruitment, training or employment strategies. Where they had thought about the problem, the most common response was to try to increase the attractiveness of their jobs through improved liaison with schools and colleges, better marketing of vacancies and improved training and career prospects, the latter being seen as more relevant than increasing wages. Unfortunately, these strategies simply increase market share in a shrinking market. But the emphasis being given to training augurs well for the long-term development of the United Kingdom's skill base. Another important and desirable strategy is the substitution of YTS entrants for more highly qualified recruits. But many of these companies are already having problems filling their existing YTS places and they will have to improve the attractiveness of their schemes to maintain intakes, let alone to increase them.

Among other options, the importance of retraining is increasingly being recog-

nized but was not expected to make a significant contribution. A few large organizations have started reducing entry standards or introducing aptitude testing in place of a qualification-based selection. Few, however, had started looking to new sources of recruits such as the married women returners, the unemployed, mature entrants or other sources. Finally, in a few instances, companies were introducing new technology or productivity improvement programmes to reduce the demand for staff, or relocating to easier labour markets, although such action was usually prompted by other business considerations.

As awareness of the problem grows, the most likely response is to try to increase

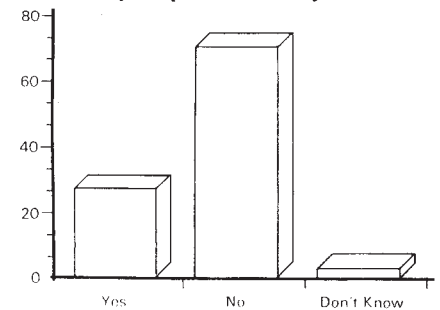


Fig. 2 Few employers answer "yes" when asked "Are you considering revising recruitment policies?".

market share. All this presupposes that nothing else is happening. But in the most pressured part of the market, those with five or more O levels, or two A levels, additional pressure is coming from the education system to get more pupils to stay in education longer.

Who then will win the battle for the bright school-leaver with numerate O and A levels? The banks offering high salaries, good training and career prospects, or the engineering courses in polytechnics offering good long-term career prospects, but initially small grants or loans? The clear losers will be organizations unable or unwilling to adapt their recruitment strategies and reward structures in the light of increased competition. The winners will be the young, particularly those with good skills and motivation who are prepared to follow the jobs and opportunities to those parts of the economy where they exist. □

Richard Pearson is at the Institute of Manpower Studies, Mantell Building, University of Sussex, Brighton BN1 9RF, UK.